

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2003, 08:35:50 ; Search time 66 Seconds
(without alignments)
205.933 Million cell updates/sec

Title: US-09-882-434A-1
Perfect score: 551
Sequence: 1 MASTKLFPSVITVMMLAMA.....FGSARACNPGKMSITQC 102.

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: /SID2/gcgdata/geneeq/geneeq-emb1/AA1980.DAT.*
2: /SID2/gcgdata/geneeq/geneeq-emb1/AA1981.DAT.*
3: /SID2/gcgdata/geneeq/geneeq-emb1/AA1982.DAT.*
4: /SID2/gcgdata/geneeq/geneeq-emb1/AA1983.DAT.*
5: /SID2/gcgdata/geneeq/geneeq-emb1/AA1984.DAT.*
6: /SID2/gcgdata/geneeq/geneeq-emb1/AA1985.DAT.*
7: /SID2/gcgdata/geneeq/geneeq-emb1/AA1986.DAT.*
8: /SID2/gcgdata/geneeq/geneeq-emb1/AA1987.DAT.*
9: /SID2/gcgdata/geneeq/geneeq-emb1/AA1988.DAT.*
10: /SID2/gcgdata/geneeq/geneeq-emb1/AA1989.DAT.*
11: /SID2/gcgdata/geneeq/geneeq-emb1/AA1990.DAT.*
12: /SID2/gcgdata/geneeq/geneeq-emb1/AA1991.DAT.*
13: /SID2/gcgdata/geneeq/geneeq-emb1/AA1992.DAT.*
14: /SID2/gcgdata/geneeq/geneeq-emb1/AA1993.DAT.*
15: /SID2/gcgdata/geneeq/geneeq-emb1/AA1994.DAT.*
16: /SID2/gcgdata/geneeq/geneeq-emb1/AA1995.DAT.*
17: /SID2/gcgdata/geneeq/geneeq-emb1/AA1996.DAT.*
18: /SID2/gcgdata/geneeq/geneeq-emb1/AA1997.DAT.*
19: /SID2/gcgdata/geneeq/geneeq-emb1/AA1998.DAT.*
20: /SID2/gcgdata/geneeq/geneeq-emb1/AA1999.DAT.*
21: /SID2/gcgdata/geneeq/geneeq-emb1/AA2000.DAT.*
22: /SID2/gcgdata/geneeq/geneeq-emb1/AA2001.DAT.*
23: /SID2/gcgdata/geneeq/geneeq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	551	100.0	102	18	AAW31738
2	74.5	13.5	486	21	AAAB10456
3	71	12.9	1280	22	AAAB1064
4	71	12.9	1283	22	AAAB1065
5	70.5	12.8	267	22	AAAB96121
6	70.5	12.8	267	22	AAAB96431
7	70.5	12.8	267	22	AAAB17539
8	70.5	12.8	524	19	AAW3498
9	69	12.5	138	22	AAU50762
10	69	12.5	1280	22	AAAB1068

11	69	12.5	1280	22	AAE00304	Dog P-glycoprotein
12	69	12.5	1281	22	AAE00303	Dog (FGP) P-glycop
13	69	12.5	1281	22	AAE00308	Dog P-glycoprotein
14	69	12.5	1281	22	AAE00309	Dog P-glycoprotein
15	69	12.5	1281	22	AAE00310	Dog P-glycoprotein
16	68.5	12.4	1280	19	AAW48998	N-terminal single
17	68.5	12.4	1280	19	AAW49000	Human P glycoprote
18	68	12.3	1280	8	AAW70452	Sequence encoded b
19	68	12.3	1280	11	AAW04868	Protein encoded by
20	68	12.3	1280	14	AAW44297	Sequence encoded b
21	68	12.3	1280	15	AAW63524	Human P-glycoprote
22	68	12.3	1280	16	AAW44073	Human P-glycoprote
23	68	12.3	1280	19	AAW48997	Human multidiug re
24	68	12.3	1280	19	AAW48999	wild-type human P
25	68	12.3	1280	21	AAW58186	C-terminal single
26	68	12.3	1280	21	AAW58187	Human wild-type mu
27	68	12.3	1280	22	AAU04347	Human G155V mutant
28	68	12.3	1280	22	AAW81066	Human multidiug re
29	68	12.3	1280	22	AAW81959	Human P-glycoprote
30	68	12.3	1280	22	AAE00306	Human MDRI. Homo
31	68	12.3	1280	23	AAE24211	Human P-glycoprote
32	68	12.3	1280	23	AAU97493	Human MDR protein.
33	68	12.3	1280	23	AAE18964	Human P-glycoprote
34	68	12.3	1280	23	AAE18965	Human wild type P-
35	68	12.3	1280	23	AAE18966	Human P-gp mutant,
36	68	12.3	1280	23	AAE18967	Human P-gp mutant,
37	68	12.3	1280	23	AAE18968	Human P-gp mutant,
38	68	12.3	1280	23	AAE18969	Human P-gp mutant,
39	68	12.3	1280	23	AAE18970	Human P-gp mutant,
40	68	12.3	1280	23	AAE18971	Human P-gp mutant,
41	68	12.3	1280	23	AAE18972	Human P-gp mutant,
42	68	12.3	1280	23	AAE18973	Human P-gp mutant,
43	68	12.3	1280	23	AAE18974	Human P-gp mutant,
44	68	12.3	1280	23	AAE18975	Human P-gp mutant,
45	68	12.3	1280	23	AAE18976	Human P-gp mutant,

ALIGNMENTS

RESULT 1	AAW31738	standard; Protein; 102 AA.
ID	AAW31738	
XX		
AC	AAW31738;	
XX		
DT	27-MAR-1998	(first entry)
XX		
DE	Macadamia integrifolia antimicrobial protein 1.	
XX		
KW	Antimicrobial protein 1; growth inhibition; fungus; bacterium;	
KW	fungi; bacteria; pathogen control.	
XX		
OS	Macadamia integrifolia.	
XX		
FT	Key	Location/Qualifiers
FT	Peptide	1..26
FT		/label= sig_peptide
FT	Peptide	27..102
FT		/label= mat_peptide
XX		
PN	WO9728185-A1.	
XX		
PD	07-AUG-1997.	
XX		
PF	31-JAN-1997;	97WO-AU00052.
XX		
PR	31-JAN-1996;	96AU-0007802.
XX		
PA	(RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.	
XX		
PI	Goulter KC, Green JL, Harrison SJ, Manners JM, Marcus JP;	
XX		

DR WPI; 1997-448317/41.
DR N-PSDB; AAT88851.
XX
PT New anti-microbial protein from *Macadamia integrifolia* - active in
PT inhibiting the growth of fungi and bacteria in plants and animals
XX
PS Claim 1; Pages 23-24; 38pp; English.
XX
CC The present sequence is *Macadamia integrifolia* antimicrobial
CC protein 1, which exerts inhibitory activity on the growth of fungi
CC and bacteria. It can be used for the control of pathogens in plants
CC and animals.
XX
SQ Sequence 102 AA;
Query Match 100.0%; Score 551; DB 18; Length 102;
Best Local Similarity 100.0%; Pred. No. 1.5e-54;
Matches 102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MASTKLFSSVITVNMILAMASEMVGSAFTWSPGCCNNRAERYSKGCCSAIHQKGGYDF 60
DB 1 MASTKLFSSVITVNMILAMASEMVGSAFTWSPGCCNNRAERYSKGCCSAIHQKGGYDF 60
QY 61 SYTGTAALYNQAGCSGVAHTRFGSSARACNPFQGWKSGIFIQ 102
DB 61 SYTGTAALYNQAGCSGVAHTRFGSSARACNPFQGWKSGIFIQ 102
RESULT 2
AAB10456
ID AAB10456 standard; Protein; 486 AA.
XX
AC AAB10456;
XX
DT 11-DEC-2000 (first entry)
XX
DE S. pastorianus mutant hexokinase protein.
XX
KW Hexokinase; mutant; creatinine kinase detection; glucose detection.
XX
OS Saccharomyces pastorianus.
XX
PN JP2000175688-A.
XX
PD 27-JUN-2000.
XX
PF 17-DEC-1998; 98JP-0359018.
XX
PR 17-DEC-1998; 98JP-0359018.
XX
PS (TOYM) TOYOBO KK.
XX
WPI; 2000-501189/45.
DR N-PSDB; AAA71483.
XX
PT Mutant hexokinase, useful for the detection of creatinine kinase and
PT glucose, comprises a deletion, addition or insertion in the wild-type
PT amino acid sequence.
XX
PS Claim 3; Page 8-9; 13pp; Japanese.
XX
CC This invention describes a novel mutant hexokinase (I) comprising a
CC protein having hexokinase activity in which at least one amino acid
CC is added, deleted, inserted or replaced. The amino acid sequence
CC constitutes the protein having hexokinase activity which remains after
CC the protein is treated at 50 degrees C for 30 minutes in liquid state,
CC and which is higher than that of the protein before it is mutated. (I)
CC is used for the detection of creatinine kinase and glucose. This sequence
CC represents the *Saccharomyces pastorianus* hexokinase protein which is
CC described in the method of the invention.
XX
SQ Sequence 486 AA;

Query Match 13.5%; Score 74.5; DB 21; Length 486;
Best Local Similarity 29.2%; Pred. No. 5.5;
Matches 28; Conservative 10; Mismatches 27; Indels 31; Gaps 6;
QY 8 FSVITVMM---LIAMASEMVGSAFTWSPGCCNNRAERYSKGCCSAIHQKGGYDFSYTG 64
DB 369 FGINTTVQERKLIRRLSELIGA-----RAARLSVCGIAAIVCQKRGYK---TG 412
QY 65 QTAALYNQAGCSGVAHTRFF-CSSARACNP-----FGW 95
DB 413 HIAA-----DGSVYNRYPGFKERAAALAKDIYGW 441
RESULT 3
AAB81064
ID AAB81064 standard; Protein; 1280 AA.
XX
AC AAB81064;
XX
DT 25-JUN-2001 (first entry)
XX
DE Cynomologous monkey P-glycoprotein variant 1.
XX
KW Cynomologous monkey; P-glycoprotein; PGP; multidrug transporter; MDRL;
KW efflux pump.
XX
OS Macaca fascicularis.
XX
PN WO200123565-A1.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26592.
XX
PR 28-SEP-1999; 99US-0156921.
XX
PR 12-OCT-1999; 99US-0158818.
XX
PA (GENT-) GENTEST CORP.
XX
PI Stocker PJ, Steimel-Crespi DT, Crespi CL;
XX
WPI; 2001-316136/33.
DR N-PSDB; AAF86127.
XX
PT Novel isolated nucleic acid encoding cynomologous monkey P-glycoprotein
PT (PGP) and homologous PGP polypeptides are useful for predicting
PT bioavailability of compound and increasing PGP transporter activity in
PT cell.
XX
PS Claim 9; Page 57-59; 84pp; English.
XX
CC This invention relates to a polynucleotide sequence encoding a
CC cynomologous monkey P-glycoprotein (PGP), and an allelic variant of the
CC PGP protein. PGP, also known as multidrug transporter, MDRL is a member
CC of the ABC transporter superfamily. The enzyme serves as an efflux pump
CC exporting small molecules across the cell membrane. The invention
CC includes a cynomologous monkey (Macaca fascicularis) PGP coding sequence
CC and protein, and also that of an allelic variant. The PGP polynucleotide
CC sequence is useful for increasing PGP transporter activity in a cell.
CC Antisense sequences of the cDNA are useful for inhibiting PGP transport
CC activity in a mammalian cell. They may also be used for increasing the
CC bioavailability of a drug. The present sequence represents the
CC cynomologous monkey P-glycoprotein.
XX
SQ Sequence 1280 AA;
Query Match 12.9%; Score 71; DB 22; Length 1280;
Best Local Similarity 25.5%; Pred. No. 42;
Matches 25; Conservative 12; Mismatches 33; Indels 28; Gaps 3;
QY 6 LFPSVITVMMILAMASEMV-----NGSAFTWSPGCCNNRAERYSKGCCSAIHQKGGYD 59
DB 334 VFPSVLIGAFSVGQASPSIEAFANARGAFAEIKIDNKPISDSYSKSGHKPDNTKGNLE 393

OY 60 -----FSY-----TGCTAALYNQAGC 75
DB 394 FRNVHFSYPSRKEVKILKGLNLKVGSGQTVALVGNNGC 434

RESULT 4

ABAB1065
ID AAB81065 standard; Protein; 1283 AA.

AC AAB81065;

DT 25-JUN-2001 (first entry)

DE Cynomologous monkey P-glycoprotein variant 2.

KM Cynomologous monkey; P-glycoprotein; PGP; multidrug transporter; MDR1;
KW efflux pump.

OS Macaca fascicularis.

XX Key Location/Qualifiers

FT Misc-difference 93..95
FT to PGP variant AAB81064"

XX MO200123565-A1.

XX 05-APR-2001.

XX 28-SEP-2000; 2000WO-US26592.

XX 28-SEP-1999; 99US-0156921.

XX 12-OCT-1999; 99US-0158818.

XX (GENT-) GENTEST CORP.

XX Stocker PJ, Steimel-Crepi DT, Crepi CL;

XX WPI; 2001-316136/33.

XX N-PSDB; AAF86128.

XX Novel isolated nucleic acid encoding cynomologous monkey P-glycoprotein
(PGP) and homologous PGP polypeptides are useful for predicting
PT bioavailability of compound and increasing PGP transporter activity in
PT cell.

XX Claim 9; Page 65-68; 84pp; English.

XX This invention relates to a polynucleotide sequence encoding a
XX cynomologous monkey P-glycoprotein (PGP), and an allelic variant of the
XX PGP protein. PGP, also known as multidrug transporter, MDR1 is a member
XX of the ABC transporter superfamily. The enzyme serves as an efflux pump
XX exporting small molecules across the cell membrane. The invention
XX includes a cynomologous monkey (Macaca fascicularis) PGP coding sequence
XX and protein, and also that of an allelic variant. The PGP polynucleotide
XX sequence is useful for increasing PGP transporter activity in a cell.
XX Antisense sequences of the cDNA are useful for inhibiting PGP transport
XX activity in a mammalian cell. They may also be used for increasing the
XX bioavailability of a drug. The present sequence represents the
XX cynomologous monkey P-glycoprotein variant 2. The protein has an
XX additional 3 amino acids when compared to PGP variant 1 (AAB81065).

XX Sequence 1283 AA;

Query Match 12.9%; Score 71; DB 22; Length 1283;

Best Local Similarity 25.5%; Pred. No. 42;

Matches 25; Conservative 12; Mismatches 33; Indels 26; Gaps 3;

OY 6 LFEEVITVMMILIMASENV-----NSAFTVMSGPCNNRERYSKGCSAIIHQKGYD 59
DB 337 VFESVLIGAFVQASPSIEAFANARGAAFEIFKIIDNKPSIDSKSGHKHPINIKNLE 396

OY 60 -----FSY-----TGCTAALYNQAGC 75
DB 397 FRNVHFSYPSRKEVKILKGLNLKVGSGQTVALVGNNGC 434

RESULT 5

ABAB96121
ID ABB96121 standard; Protein; 267 AA.

AC ABB96121;

DT 21-JUN-2002 (first entry)

DE Human testicular antigen SEQ ID NO: 1505.

KM Human; testicular antigen; testes; cancer; metastasis; immune disorder;
KW reproductive system disorder; urinary system disorder; gene therapy;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disease; infection; cytostatic.

XX Homo sapiens.

XX MO200155317-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01329.

XX 31-JAN-2000; 2000US-0179065.

XX 04-FEB-2000; 2000US-0180628.

XX 24-FEB-2000; 2000US-0184664.

XX 02-MAR-2000; 2000US-0186350.

XX 16-MAR-2000; 2000US-0189874.

XX 17-MAR-2000; 2000US-0190076.

XX 18-APR-2000; 2000US-0198123.

XX 19-MAY-2000; 2000US-0205515.

XX 07-JUN-2000; 2000US-0209467.

XX 28-JUN-2000; 2000US-0214886.

XX 30-JUN-2000; 2000US-0215135.

XX 07-JUL-2000; 2000US-0216647.

XX 07-JUL-2000; 2000US-0216880.

XX 11-JUL-2000; 2000US-0217487.

XX 11-JUL-2000; 2000US-0217496.

XX 14-JUL-2000; 2000US-0218290.

XX 26-JUL-2000; 2000US-0220963.

XX 26-JUL-2000; 2000US-0220964.

XX 14-AUG-2000; 2000US-0224518.

XX 14-AUG-2000; 2000US-0224519.

XX 14-AUG-2000; 2000US-0225213.

XX 14-AUG-2000; 2000US-0225214.

XX 14-AUG-2000; 2000US-0225266.

XX 14-AUG-2000; 2000US-0225267.

XX 14-AUG-2000; 2000US-0225268.

XX 14-AUG-2000; 2000US-0225270.

XX 14-AUG-2000; 2000US-0225270.

XX 14-AUG-2000; 2000US-0225447.

XX 14-AUG-2000; 2000US-0225757.

XX 14-AUG-2000; 2000US-0225758.

XX 14-AUG-2000; 2000US-0225759.

XX 18-AUG-2000; 2000US-0226279.

XX 22-AUG-2000; 2000US-0226281.

XX 22-AUG-2000; 2000US-0226286.

XX 22-AUG-2000; 2000US-0226286.

XX 22-AUG-2000; 2000US-0227182.

XX 23-AUG-2000; 2000US-0227182.

XX 30-SEP-2000; 2000US-0228924.

XX 01-SEP-2000; 2000US-0229287.

XX 01-SEP-2000; 2000US-0229344.

XX 01-SEP-2000; 2000US-0229345.

XX 05-SEP-2000; 2000US-0229509.

XX 05-SEP-2000; 2000US-0229513.

XX 06-SEP-2000; 2000US-0230437.

XX 06-SEP-2000; 2000US-0230438.

XX 08-SEP-2000; 2000US-0231242.

PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231245.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 23-SEP-2000; 2000US-0234597.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 29-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 13-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.

PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 01-DEC-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-483232/52.
 XX Nucleic acids encoding 973 human testicular antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating testicular cancer -
 XX Claim 11; SEQ ID NO 1505; 766pp; English.
 XX The present invention provides the protein and coding sequences of 973
 CC human testicular antigens, and fragments of their genomic sequences. The
 CC sequences can be used in the treatment of cardiovascular, urinary system,
 CC reproductive system, immune, respiratory, neurological and
 CC gastrointestinal disorders, infections, and particularly cancer,
 CC especially testicular cancers. The present sequence is a protein of the
 CC invention.
 XX Sequence 267 AA;
 SQ
 Query Match 12.8%; Score 70.5; DB 22; Length 267;
 Best Local Similarity 27.0%; Pred. No. 7.9;
 Matches 20; Conservative 13; Mismatches 36; Indels 5; Gaps 2;
 QY 28 APTVSGPCNNRARRYSKCGSAHQKGGYDFSYTGQTAALYNQAGCGVAHTFGSSA 87
 DB 146 SFEFMDSPFNSDRGRGH--GLRGAFSAFGFEPFAFMFAFSFNLGCGSGSHHTFSSTS 203
 QY 88 --RACNPFQWKS I 98
 DB 204 FGGSSGSGSGFKSV 217
 RESULT 6
 AAM95431
 ID AAM95431 standard; Protein; 267 AA.
 XX AC AAM95431;
 XX DT 21-NOV-2001 (first entry)
 XX DE Human reproductive system related antigen SEQ ID NO: 4089.
 XX KW Human; reproductive system related antigen; reproductive system disorder;
 XX OS cancer; gene therapy.
 XX OS Homo sapiens.
 XX PN WO200155320-A2.
 XX PD 02-AUG-2001.
 XX PF 17-JAN-2001; 2001WO-US01339.

XX 31-JAN-2000; 2000US-0179055.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225256.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226686.
 PR 23-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227189.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 29-SEP-2000; 2000US-0236370.

PR 02-OCT-2000; 2000US-0235802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251889.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.

PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM,
 XX
 DR WPI; 2001-46570/50.
 DR N-PSDB; AAL01401.
 XX
 PT Isolated nucleic acid molecule encoding a reproductive system antigen

PT is used in preventing, treating or ameliorating a medical condition -
XX Claim 11; SEQ ID NO 4089; 1297pp + Sequence Listing; English.
XX
XX The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a protein of the invention.
XX
SQ Sequence 267 AA;
Query Match 12.8%; Score 70.5; DB 22; Length 267;
Best Local Similarity 27.08; Pred. No. 7.9;
Matches 20; Conservative 13; Mismatches 36; Indels 5; Gaps 2;
QY 28 AFTVMSGPCNNRERYSKCCGCAHQGGYDFSYTGQTAALYNOAGCGVHAHTRFGSSA 87
Db 146 SFPEWDSFNSDGRGRGH-GLRGAFSAAGFGFPAFMEAFSSFNMLGCGSGSHTTFSST 203
QY 88 ---RACNFFGHWKSI 98
Db 204 FGGSSSGSGPKSV 217
RESULT 7
AAU17599
ID AAU17599 standard; Protein; 267 AA.
XX
AC AAU17599;
XX
DT 07-NOV-2001 (first entry)
XX
DE Novel signal transduction pathway protein, Seq ID 1164.
XX
KW Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;
KW anti-inflammatory; anti-HIV; antibacterial; anti-inflammatory; cancer;
KW immune system disorder; rheumatoid arthritis; inflammatory condition;
KW organ transplant rejection; infection; hepatitis C; blood disorder;
KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;
KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;
KW reproductive system; gastrointestinal; liver disorder; AIDS;
KW acquired immune deficiency syndrome.
XX
XX Homo sapiens.
OS
XX
PN WO200154733-A1.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01312.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 08-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.

PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249254.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI: 2001-465460/50.
XX
XX N-PSDB; AAS27516.
XX
XX Novel polypeptides useful for diagnosing, treating, preventing and/or
XX prognosing disorders related to the proteins, including cancers, immune
XX disorders and neuronal disorders -
XX
PS Claim 1; SEQ ID No 1164; 880pp; English.

CC disorder (inflammatory disorders), liver disorders (cirrhosis),
CC as stimulators of B-cell responsiveness to pathogens, activators of
CC T-cells, to induce higher affinity antibodies, and as a means to induce
CC tumour proliferation in pathologies e.g. acquired immune deficiency
CC syndrome (AIDS). AAU7059-AAU17683 represent novel signal transduction
CC pathway protein, amino acid sequences of the invention.
XX
Query Match 12.8%; Score 70.5; DB 22; Length 267;
Best Local Similarity 27.0%; Pred No. 7.9;
Matches 20; Conservative 13; Mismatches 36; Indels 5; Gaps 2;
QY 28 AFTVMSGPCNNRERYSKCGSAHQKGYDPSYTGTAALYNAGCGSVATFRGSSA 87
DB 146 SFEFWDSPFNSDRGRGH--GLRFAFSAGFGEPPAPMEAFSSFMMLGCGSGHTTFSST 203
QY 88 ---RACNPFQWMSI 98
DB 204 FGGSSGSSGSGFRSV 217
RESULT 8
AAW34998
ID AAW34998 standard; Protein; 524 AA.
XX
XX AAW34998;
XX
XX AC
XX
XX 21-MAY-1998 (first entry)
XX
XX DT
XX
XX DE Archaeobacterium AEP11a endoglucanase.
XX
XX KW Endoglucanase; cellulase; carboxymethylcellulose; cellulose;
XX biomass; beta-1,4-glycosidic bond; hydrolase; saccharification;
XX KM thermostable enzyme; thermophilic; glycosidase.
XX
XX OS Archaeobacterium isolate AEP11a (Clone 63GP1).
XX
XX PN WO9744361-A1.
XX
XX PD 27-NOV-1997.
XX
XX PF 22-MAY-1997; 97WO-US08793.
XX
XX PR 22-MAY-1996; 96US-0651572.
XX
XX PA (RECO-) RECOMBINANT BIOCATALYSIS INC.
XX
XX PI Lam DE, Mathur EJ;
XX
XX WPI: 1998-018435/02.
XX
XX DR N-PSDB; AAT94206.
XX
XX PT Endoglucanase(s), preferably from archaeal bacterium, AEP11a -
XX useful to degrade carboxymethylcellulose and hydrolyse of
XX beta-1,4-glycosidic bonds in cellulose
XX
PS Claim 1; Fig 1N; 164pp; English.

This protein comprises an endoglucanase of archaeobacterium AEP11a (Clone 63GP1), a hydrothermal vent isolate. The endoglucanase is capable of degrading carboxymethylcellulose and of hydrolysing the beta-1,4-glycosidic bonds in cellulose. It has homology to another endoglucanase (see AAW34985) of archaeobacterium AEP11a. It can be produced from native cells or from recombinant host cells, especially prokaryotic host cells transformed with a plasmid or virus-derived vector including the endoglucanase DNA (see AAT94206). 24 Endoglucanases (see AAW34986-W35008) are claimed. They can be used to degrade cellulose for the conversion of plant biomass into fuels and chemicals, for use in detergents, textiles, animal feed, waste treatment, and in the fruit juice and brewing industries for the clarification and extraction of juices.

Sequence 524 AA:
SQ

Query Match 12.8%; Score 70.5; DB 19; Length 524;
 Best Local Similarity 26.4%; Pred. No. 17;
 Matches 33; Conservative 12; Mismatches 37; Indels 43; Gaps 8;

QY 3 STKLFFSV-ITVMMILMA-----SEMVNGSAFT---VMSGPG-----CNRR 40
 DB 63 SMKKFVALFITMFFVSMVAQPAASAAKYSELEGGVIMQAFYDVPGGGIWMDTIRSK 122
 QY 41 AERYSKGCGSAI-----HOKGG-----YDFSVTGTAAALYNQAGCGVAHTRFGSS 86
 DB 123 IPEYEAIGSAIWIPIPPASKMGSGYSGYDYPDFDLGE-----YNO---XGTYETRFSGK 175
 QY 87 ARACN 91
 DB 176 QELIN 180

RESULT 9
 ID AAU50762 standard; Protein; 138 AA.
 AC AAU50762;
 DT 27-FEB-2002 (first entry)
 DE Propionibacterium acnes immunogenic protein #11658.
 KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KW dermatological; osteopathic; neuroprotectant.
 OS Propionibacterium acnes.
 PN W0200181581-A2.
 XX 01-NOV-2001.
 XX 20-APR-2001; 2001WO-US12865.
 XX 21-APR-2000; 2000US-199047P.
 PR 02-JUN-2000; 2000US-208841P.
 PR 07-JUL-2000; 2000US-216747P.
 XX (CORI-) CORIXA CORP.
 XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
 PI L'malsonneuve J, Zhang Y, Jen S, Carter D;
 XX WPI; 2001-616774/71.
 DR N-PSDB; AAS59549.
 XX Propionibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris -
 XX Example 1; SEQ ID No 11957; 1069pp; English.
 XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
 CC polypeptides. The proteins and their associated DNA sequences are used in
 CC the treatment, prevention and diagnosis of medical conditions caused by
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
 CC P. acnes is also involved in infections of bone, joints and the central
 CC nervous system, however it is particularly involved in the inflammatory
 CC lesions associated with acne vulgaris. A method for detecting the
 CC presence or absence of P. acnes in a patient comprises contacting a
 CC sample with a binding agent that binds to the proteins of the invention
 CC and determining the amount of bound protein in the sample. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC specific for P. acnes proteins. These antibodies can be used to
 CC downregulate expression and activity of P. acnes polypeptides and

CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA).
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 138 AA;
 SQ

Query Match 12.5%; Score 69; DB 22; Length 138;
 Best Local Similarity 31.5%; Pred. No. 5.4;
 Matches 29; Conservative 8; Mismatches 25; Indels 30; Gaps 7;

QY 14 MMLTAMASEMVNGSAFTVW-SGPCNNRAERYSKGCSAIHQKGYDFSVTGTAAALYNQ 72
 DB 53 LMLRLPRESVQ-----WPSGPASIFRA-----SCGTST-----TASTSAAM-V 90
 QY 73 AGCGVAHTRFGS-----SARACNPFQWKS 98
 DB 91 GSCSRIETAGAAASSPRISRLASP-GWKS 121

RESULT 10
 AAB81068
 ID AAB81068 standard; Protein; 1280 AA.
 XX AC AAB81068;
 XX 25-JUN-2001 (first entry)
 DT Dog P-glycoprotein SEQ ID 7.
 DE Cynomologous monkey; P-glycoprotein; PGP; multidrug transporter; MDR1;
 DE efflux pump; dog.
 KW Canis familiaris.
 OS WO200123565-A1.
 XX 05-APR-2001.
 XX 28-SEP-2000; 2000WO-US26592.
 XX 28-SEP-1999; 99US-0156921.
 PR 12-OCT-1999; 99US-0158818.
 XX (GENT-) GENTEST CORP.
 XX Stocker PJ, Steimel-Crespi DT, Crespi CL;
 XX WPI; 2001-316136/33.
 XX Novel isolated nucleic acid encoding cynomologous monkey P-glycoprotein
 PT (PGP) and homologous PGP polypeptides are useful for predicting
 PT bioavailability of compound and increasing PGP transporter activity in
 PT cell -
 XX Claim 8; Page 73-76; 84pp; English.
 PS This invention relates to a polynucleotide sequence encoding a
 CC cynomologous monkey P-glycoprotein (PGP), and an allelic variant of the
 CC PGP protein. PGP, also known as multidrug transporter, MDR1 is a member
 CC of the ABC transporter superfamily. The enzyme serves as an efflux pump
 CC exporting small molecules across the cell membrane. The invention
 CC includes a cynomologous monkey (Macaca fascicularis) PGP coding sequence
 CC and protein, and also that of an allelic variant. The PGP polynucleotide
 CC sequence is useful for increasing PGP transporter activity in a cell.
 CC Antisense sequences of the cDNA are useful for inhibiting PGP transport
 CC activity in a mammalian cell. They may also be used for increasing the
 CC bioavailability of a drug. The present sequence represents
 CC P-glycoprotein from a dog.
 XX Sequence 1280 AA;
 SQ

QY 6 LFFSVITVWMLIAMSEV-----NSAFTVWSPGCCNNRERYSKGCSAIIHOKGYD 59
 DB 336 VFESVLIGAFSIGQASPSIEAFANARGAAVEIFKIIDNKPSIDSYSGHKRPDNIKNLE 395
 QY 60 -----FSY-----TGQTALYNQAGC 75
 DB 396 FKXVHFSPSRKEVKILKGLNKVQSGQTVALVGNSSC 433

RESULT 15
 ID AAE00310 standard; Protein; 1281 AA.
 AC AAE00310;
 DT 13-JUN-2001 (first entry)
 XX Dog P-glycoprotein (PGP) allelic variant (Genotype D).
 DE Dog; P-glycoprotein (PGP) allelic variant; PGP, multidrug transporter;
 KW MDR1; drug bioavailability; transgenic animal; genetic model.
 XX Canis familiaris.
 OS
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 25 /note= "Asn of GenotypeC substituted by Lys"
 FT Misc-difference 197 /note= "His of GenotypeC substituted by Gln"
 FT Misc-difference 329 /note= "Ser of GenotypeC substituted by Thr"
 FT Misc-difference 1148 /note= "Met of GenotypeC substituted by Val"
 PN WQ200123540-A2.
 XX
 PD 05-APR-2001.
 XX
 PF 28-SEP-2000; 2000WO-US26767.
 XX
 PR 28-SEP-1999; 99US-0156510.
 XX
 PA (GENT-) GENTEST CORP.
 PI Stocker PJ, Steimel-crespi DT, Crespi CL, Reif TC, Patten CJ;
 XX WPI, 2001-235373/24.
 DR N-PSDB; AAD03506.
 XX
 PT New dog P-glycoproteins (PGP) and their encoding nucleic acids, useful
 PT for determining the bioavailability of drugs and for screening for dog
 PT PGP inhibitors -
 PS
 PS Claim 17, Page 108-110, 111pp; English.
 XX
 CC The invention relates to dog P-glycoprotein (PGP) also referred
 CC as multidrug transporter (MDR1) and nucleic acids encoding them.
 CC The invention also includes fragments and biologically functional
 CC variants of dog P-glycoprotein. PGP and their nucleic acids are
 CC useful for determining the bioavailability of drugs and for
 CC screening PGP inhibitors. They are useful for the diagnosis and
 CC treatment of conditions characterized by PGP activity, by
 CC reducing or increasing PGP activity in a cell, PGP nucleic acids
 CC are used as oligonucleotide probes. Complements of PGP nucleic
 CC acids are useful as antisense oligonucleotides, to induce a PGP
 CC "knockout" phenotype. They are used to prepare a non-human
 CC transgenic animal, which are valuable as genetic models for
 CC human diseases.
 CC The present sequence is dog P-glycoprotein (PGP) allelic variant.
 CC This sequence is also referred as Genotype D protein. The PGP
 CC enzyme functions as an efflux pump exporting small molecules
 CC across the cell membrane. This enzyme is a member of the ABC

CC transporter family.
 XX
 SQ Sequence 1281 AA;
 Query Match 12.5%; Score 69; DB 22; Length 1281;
 Best Local Similarity 25.5%; Pred. No. 70;
 Matches 25; Conservative 12; Mismatches 33; Indels 28; Gaps 3;
 QY 6 LFFSVITVWMLIAMSEV-----NSAFTVWSPGCCNNRERYSKGCSAIIHOKGYD 59
 DB 336 VFESVLIGAFSIGQASPSIEAFANARGAAVEIFKIIDNKPSIDSYSGHKRPDNIKNLE 395
 QY 60 -----FSY-----TGQTALYNQAGC 75
 DB 396 FKXVHFSPSRKEVKILKGLNKVQSGQTVALVGNSSC 433

Search completed: January 12, 2003, 09:35:53
 Job time : 68 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2003, 09:36:45 (Search time 25 Seconds)

(without alignments)
120.046 Million cell updates/sec

Title: US-09-882-434a-1

Perfect score: 551
Sequence: 1 MASTKLFVETVMMILAWA.....FGSSARACNPFQKSTFIQC 102

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

Issued Patents AA:*
1: /cgn2_6/prodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/iaa/5A_COMB.pep:*
4: /cgn2_6/prodata/1/iaa/5B_COMB.pep:*
5: /cgn2_6/prodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/1/iaa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the total being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	74.5	13.5	486	1	US-07-872-678A-48
2	68	12.3	1280	2	US-08-583-276-19
3	68	12.3	1280	4	US-08-752-447-2
4	68	12.3	1280	4	US-09-316-167-2
5	68	12.3	1280	6	5206352-4
6	67.5	12.3	902	1	US-08-701-846-2
7	63.5	11.5	1019	1	US-08-296-014A-4
8	63.5	11.5	1019	2	US-08-596-405-4
9	63.5	11.5	1019	2	US-08-877-620-4
10	63.5	11.5	1083	1	US-08-296-014A-2
11	63.5	11.5	1083	2	US-08-596-405-2
12	63.5	11.5	1083	2	US-08-877-620-2
13	62.5	11.3	135	2	US-07-857-224B-99
14	62	11.3	442	4	US-09-347-833-11
15	62	11.3	3729	2	US-08-804-227C-4
16	61.5	11.2	111	2	US-07-857-224B-105
17	61.5	11.2	135	2	US-07-857-224B-97
18	61.5	11.2	135	2	US-07-857-224B-98
19	61	11.1	510	4	US-09-212-247C-5
20	61	11.1	1196	1	US-08-144-121-4
21	61	11.1	1196	2	US-08-735-893-4
22	61	11.1	1279	2	US-08-784-649A-2
23	60.5	11.0	985	2	US-08-680-326-41
24	60.5	11.0	1349	2	US-08-612-734B-2
25	60	10.9	687	1	US-08-470-702-10
26	60	10.9	687	1	US-08-467-831-10
27	60	10.9	730	2	US-08-696-944-2

28	60	10.9	871	4	US-09-245-041-19	Sequence 19, Appl
29	60	10.9	1198	4	US-09-245-041-131	Sequence 131, App
30	60	10.9	1198	4	US-09-794-236-3	Sequence 3, Appl1
31	60	10.9	1350	4	US-09-245-041-17	Sequence 17, Appl
32	60	10.9	2787	4	US-09-245-041-15	Sequence 15, Appl
33	59.5	10.8	305	4	US-09-230-232-1	Sequence 1, Appl1
34	59.5	10.8	861	4	US-08-960-048-12	Sequence 12, Appl
35	59	10.7	352	2	US-08-483-926A-11	Sequence 11, Appl
36	59	10.7	549	4	US-09-245-041-9	Sequence 9, Appl1
37	59	10.7	687	1	US-08-204-656B-10	Sequence 10, Appl
38	59	10.7	1260	4	US-09-245-041-2	Sequence 2, Appl1
39	59	10.7	1345	2	US-08-977-767-3	Sequence 3, Appl1
40	58.5	10.6	415	4	US-09-134-001C-3957	Sequence 357, Ap
41	58.5	10.6	510	4	US-09-134-001C-4541	Sequence 4541, Ap
42	58.5	10.6	585	4	US-09-620-412C-337	Sequence 337, App
43	58.5	10.6	1404	2	US-08-400-159-2	Sequence 2, Appl1
44	58.5	10.6	1404	3	US-08-611-729A-2	Sequence 2, Appl1
45	58.5	10.6	1752	4	US-09-556-877-180	Sequence 180, App

ALIGNMENTS

RESULT 1
US-07-872-678A-48
Sequence 48, Application US/07872678A
Patent No. 5541060
GENERAL INFORMATION:
APPLICANT: Bell, Graeme, et al.
TITLE OF INVENTION: DETECTION OF EARLY-ONSET
NON-INSULIN-DEPENDENT DIABETES MELLITUS
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: Post Office Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/872,678A
FILING DATE: 22-APRIL-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coughlin, Daniel F.
REGISTRATION NUMBER: 36,111
REFERENCE/DOCKET NUMBER: ARCD016
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713-787-1400
TELEFAX: 713-789-2679
TELEX: 79-0924
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 486 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-07-872-678A-48

Query Match 13.5%, Score 74.5, DB 1, Length 486,
Best Local Similarity 39.3%, Pred. No. 1.1,
Matches 24, Conservative 4, Mismatches 18, Indels 15, Gaps 4;
OY 40 RAERYKSGCGSIHQKGGYDFSYGQTALYNQAGCGVATRF-GSSARACNP-----FG 94
DB 391 RAARLSVCGIAIICKKRGK--TGHIA-----DGSVSTRYPGRKRNALKDIY 440

```
QY 95 W 95
Db 441 W 441

RESULT 2
US-08-583-276-19
; Sequence 19, Application US/08583276
; Patent No. 5837536
; GENERAL INFORMATION:
; APPLICANT: McDonagh, Kevin T.
; APPLICANT: Nienhuis, Arthur
; APPLICANT: Tolstoshev, Paul
; TITLE OF INVENTION: IMPROVED EXPRESSION OF HUMAN
; TITLE OF INVENTION: MULTIDRUG RESISTANCE GENES AND IMPROVED
; TITLE OF INVENTION: SELECTION OF CELLS TRANSDUCED WITH SUCH GENES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan,
; ADDRESSEE: Cecchi & Stewart
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch diskette
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: DM4 V2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/583,276
; FILING DATE: 05-JAN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/332,444
; FILING DATE: 31-OCT-1994
; APPLICATION NUMBER: 07/887,712
; FILING DATE: 22-MAY-1992
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1280 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE:
; DESCRIPTION: protein
; US-08-583-276-19

Query Match 12.3%; Score 68; DB 2; Length 1280;
Best Local Similarity 24.5%; Pred. No. 21;
Matches 24; Conservative 13; Mismatches 33; Indels 28; Gaps 3;

QY 6 LFESVITVWMLIAMSEMV-----NGSAFTVWSPGCGNNRAERYSKGCSAIHOKGGYD 59
:||||: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 334 VFFSVLIGAFSVGQASPSIEAFANARGAAVEIFKIIDNKPSIDSYSGKHGKPDNIKNLE 393
:||||: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 60 -----FSY-----TGQTAALYNQAGC 75
:||||: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 394 FRNVHFSYPSRKEVKILKGLNLKVQSGQTVALVGNSGC 431
:||||: : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 3
US-08-752-447-2
; Sequence 2, Application US/08752447
; Patent No. 5994068
; GENERAL INFORMATION:
; APPLICANT: Mechetner, Eugene
; APPLICANT: Roninson, Igor B
; TITLE OF INVENTION: Methods and Reagents for Preparing and
; TITLE OF INVENTION: Using Immunological Agents Specific for P-glycoprotein
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff Ltd.
; STREET: 300 South Wacker Drive, Seventh Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/316,167
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/752,447
; FILING DATE: 15-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; US-08-752-447-2
```

```
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff Ltd.
; STREET: 300 South Wacker Drive, Seventh Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/752,447
; FILING DATE: 15-NOV-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5994088nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 95,1121
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-9808
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1280 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-752-447-2

Query Match 12.3%; Score 68; DB 2; Length 1280;
Best Local Similarity 24.5%; Pred. No. 21;
Matches 24; Conservative 13; Mismatches 33; Indels 28; Gaps 3;

QY 6 LFESVITVWMLIAMSEMV-----NGSAFTVWSPGCGNNRAERYSKGCSAIHOKGGYD 59
:||||: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 334 VFFSVLIGAFSVGQASPSIEAFANARGAAVEIFKIIDNKPSIDSYSGKHGKPDNIKNLE 393
:||||: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 60 -----FSY-----TGQTAALYNQAGC 75
:||||: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 394 FRNVHFSYPSRKEVKILKGLNLKVQSGQTVALVGNSGC 431
:||||: : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 4
US-09-316-167-2
; Sequence 2, Application US/09316167
; Patent No. 6365357
; GENERAL INFORMATION:
; APPLICANT: Mechetner, Eugene
; APPLICANT: Roninson, Igor B
; TITLE OF INVENTION: Methods and Reagents for Preparing and
; TITLE OF INVENTION: Using Immunological Agents Specific for P-glycoprotein
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff Ltd.
; STREET: 300 South Wacker Drive, Seventh Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/316,167
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/752,447
; FILING DATE: 15-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; US-09-316-167-2
```

NAME: NO. 635357nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 95,1121
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-913-0001
TELEFAX: 312-913-9808
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1280 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-316-167-2

Query Match 12.3%; Score 68; DB 4; Length 1280;
Best Local Similarity 24.5%; Pred. No. 21;
Matches 24; Conservative 13; Mismatches 33; Indels 28; Gaps 3;

OY 6 LFFSVITVWMLIAMSSENV-----NGSAFTVWSGPGCNRRAERYSKGCGSAIHQKGYD 59
DB 334 VFFSVLIGAFSVGQASPSIEAFANARGAAYEIFKIIDNKPSIDYSKSGHKPDNIKGNLE 393

OY 60 -----FSY-----TGOTAAALYNQAGC 75
DB 394 FRNVHFSYPSRKEVKILKGLNLKVGSGQTVALVGNSSC 431

RESULT 5
5206352-4
Patent No. 5206352
APPLICANT: Ronineon, Igor B.; Pastan Ira H.; Gottesman, Michael M.
TITLE OF INVENTION: COMPOSITIONS FOR CLONES CONTAINING DNA SEQUENCES ASSOCIATED WITH MULTIDRUG RESISTANCE IN HUMAN CELLS
NUMBER OF SEQUENCES: 4
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/622,836
FILING DATE: 24-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 892,575
FILING DATE: 01-AUG-1986
APPLICATION NUMBER: 845,610
FILING DATE: 28-MAR-1986
SEQ ID NO: 4
LENGTH: 1280
5206352-4

Query Match 12.3%; Score 68; DB 6; Length 1280;
Best Local Similarity 24.5%; Pred. No. 21;
Matches 24; Conservative 13; Mismatches 33; Indels 28; Gaps 3;

OY 6 LFFSVITVWMLIAMSSENV-----NGSAFTVWSGPGCNRRAERYSKGCGSAIHQKGYD 59
DB 334 VFFSVLIGAFSVGQASPSIEAFANARGAAYEIFKIIDNKPSIDYSKSGHKPDNIKGNLE 393

OY 60 -----FSY-----TGOTAAALYNQAGC 75
DB 394 FRNVHFSYPSRKEVKILKGLNLKVGSGQTVALVGNSSC 431

RESULT 6
US-08-701-846-2
Sequence 2, Application US/08701846
Patent No. 5717069
GENERAL INFORMATION:
APPLICANT: Granados, Robert R.
TITLE OF INVENTION: DNA SEQUENCE CODING FOR A POLYPEPTIDE
TITLE OF INVENTION: WHICH ENHANCES VIRUS INFECTION OF HOST INSECTS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Barnard, Brown & Michaelis
STREET: 306 E. State St., Suite 220
CITY: Ithaca,

STATE: NY.
COUNTRY: USA
ZIP: 14850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/701,846
FILING DATE: 23-AUG-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/002,743
FILING DATE: 24-AUG-1995
ATTORNEY/AGENT INFORMATION:
NAME: Michaels, Christopher A.
REGISTRATION NUMBER: 34,390
REFERENCE/DOCKET NUMBER: BTI-32
TELECOMMUNICATION INFORMATION:
TELEPHONE: (607)273-1711
TELEFAX: (607)273-2609
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 902 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-701-846-2

Query Match 12.3%; Score 67.5; DB 1; Length 902;
Best Local Similarity 28.6%; Pred. No. 16;
Matches 20; Conservative 10; Mismatches 33; Indels 7; Gaps 4;

OY 26 GSAP--TWSGPGCNRRAE--RYSKCGCGSAIHQK-GYFSTGTGOT--AALYNQAGCGV 78
DB 213 GGAAYGAFWTAAPASTVNLGELRVSPPTMMVVIHELGHAYDFVTVTRLIEIWNNSFCRI 272

OY 79 AATRFSSAR 88
DB 273 QYTMNKTKE 282

RESULT 7
US-08-296-014A-4
Sequence 4, Application US/08296014A
Patent No. 5716834
GENERAL INFORMATION:
APPLICANT: Ding, Jeak Ling
APPLICANT: Ho, Bow
TITLE OF INVENTION: The Cloned Factor C cDNA of the
TITLE OF INVENTION: Singapore Horseshoe Crab, Carcinus scaber
TITLE OF INVENTION: rotundicauda and Purification of Factor C Proenzyme
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: 8110 Gatehouse Road, Suite 500 East
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22042
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/296,014A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy, Jr., Gerald M.
REGISTRATION NUMBER: 28,977

REFERENCE/DOCKET NUMBER: 1781-105P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1019 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-296-014A-4

Query Match 11.5%; Score 63.5; DB 1; Length 1019;
Best Local Similarity 32.8%; Pred. No. 54;
Matches 19; Conservative 8; Mismatches 22; Indels 9; Gaps 3;

Qy 32 WSG--PGCNNRAERYSKGCSAIIHQKGYDF-SYTGTAALYNQAGCGVAHTRFGSS 86
Db 314 WSGQSPSCVKVADREVDGDSKAV-----DFLDDVGEVRIHCPAGCSLTAGTVMGTA 365

RESULT 8

US-08-596-405-4
Sequence 4, Application US/08596405
Patent No. 5858706

GENERAL INFORMATION:

APPLICANT: Ding, Jeak Ling

APPLICANT: Ho, Bow

TITLE OF INVENTION: The Cloned Factor C cDNA of the

TITLE OF INVENTION: Singapore Horseshoe Crab, Carcinoscorpius

TITLE OF INVENTION: rotundicauda and Purification of Factor C Proenzyme

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Birch, Stewart, Kolasch & Birch

STREET: 8110 Gatehouse Road, Suite 500 East

CITY: Falls Church

STATE: Virginia

COUNTRY: USA

ZIP: 22042

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/596,405

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Murphy, Jr., Gerald M.

REGISTRATION NUMBER: 28,977

REFERENCE/DOCKET NUMBER: 1781-105P

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 205-8000

TELEFAX: (703) 205-8050

TELEX: 248345

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 1019 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-596-405-4

Query Match 11.5%; Score 63.5; DB 2; Length 1019;

Best Local Similarity 32.8%; Pred. No. 54;

Matches 19; Conservative 8; Mismatches 22; Indels 9; Gaps 3;

Qy 32 WSG--PGCNNRAERYSKGCSAIIHQKGYDF-SYTGTAALYNQAGCGVAHTRFGSS 86
Db 314 WSGQSPSCVKVADREVDGDSKAV-----DFLDDVGEVRIHCPAGCSLTAGTVMGTA 365

RESULT 9

US-08-877-620-4
Sequence 4, Application US/08877620
Patent No. 5985590

GENERAL INFORMATION:

APPLICANT: Ding, Jeak Ling

APPLICANT: Ho, Bow

TITLE OF INVENTION: The Cloned Factor C cDNA of the

TITLE OF INVENTION: Singapore Horseshoe Crab, Carcinoscorpius

TITLE OF INVENTION: rotundicauda and Purification of Factor C Proenzyme

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Birch, Stewart, Kolasch & Birch

STREET: 8110 Gatehouse Road, Suite 500 East

CITY: Falls Church

STATE: Virginia

COUNTRY: USA

ZIP: 22042

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/877,620

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/596,405

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Murphy, Jr., Gerald M.

REGISTRATION NUMBER: 28,977

REFERENCE/DOCKET NUMBER: 1781-105P

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 205-8000

TELEFAX: (703) 205-8050

TELEX: 248345

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 1019 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-877-620-4

Query Match 11.5%; Score 63.5; DB 2; Length 1019;

Best Local Similarity 32.8%; Pred. No. 54;

Matches 19; Conservative 8; Mismatches 22; Indels 9; Gaps 3;

Qy 32 WSG--PGCNNRAERYSKGCSAIIHQKGYDF-SYTGTAALYNQAGCGVAHTRFGSS 86
Db 314 WSGQSPSCVKVADREVDGDSKAV-----DFLDDVGEVRIHCPAGCSLTAGTVMGTA 365

RESULT 10

US-08-296-014A-2

Sequence 2, Application US/08296014A

Patent No. 5716834

GENERAL INFORMATION:

APPLICANT: Ding, Jeak Ling

APPLICANT: Ho, Bow

TITLE OF INVENTION: The Cloned Factor C cDNA of the

TITLE OF INVENTION: Singapore Horseshoe Crab, Carcinoscorpius

TITLE OF INVENTION: rotundicauda and Purification of Factor C Proenzyme

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Birch, Stewart, Kolasch & Birch

STREET: 8110 Gatehouse Road, Suite 500 East

CITY: Falls Church

STATE: Virginia

COUNTRY: USA

ZIP: 22042
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/296,014A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy, Jr., Gerald M.
REGISTRATION NUMBER: 28,977
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1083 amino acids
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-296-014A-2

Query Match 11.5%; Score 63.5; DB 1; Length 1083;
Best Local Similarity 32.8%; Pred. No. 58;
Matches 19; Conservative 8; Mismatches 22; Indels 9; Gaps 3;

QY 32 MSG--PGCNRAERYSKGCSAIHQKGYDF-STYGTALYNQAGSGVAHTRFGSS 86
DB 378 MSGQPSQVCKVADREVDCSKAV-----DFLDVGEPVRIHCPAGCSLTGTWGTGA 429

RESULT 11
US-08-596-405-2
Sequence 2, Application US/08596405
Patent No. 5858706
GENERAL INFORMATION:
APPLICANT: Ding, Jeak Ling
TITLE OF INVENTION: The Cloned Factor C cDNA of the
TITLE OF INVENTION: Singapore Horseshoe Crab, Carcinoscopus
TITLE OF INVENTION: rotundicauda and Purification of Factor C Proenzyme
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolaesch & Birch
STREET: 8110 Gatehouse Road, Suite 500 East
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22042
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/596,405
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy, Jr., Gerald M.
REGISTRATION NUMBER: 28,977
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1083 amino acids

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-596-405-2

Query Match 11.5%; Score 63.5; DB 2; Length 1083;
Best Local Similarity 32.8%; Pred. No. 58;
Matches 19; Conservative 8; Mismatches 22; Indels 9; Gaps 3;

QY 32 MSG--PGCNRAERYSKGCSAIHQKGYDF-STYGTALYNQAGSGVAHTRFGSS 86
DB 378 MSGQPSQVCKVADREVDCSKAV-----DFLDVGEPVRIHCPAGCSLTGTWGTGA 429

RESULT 12
US-08-877-620-2
Sequence 2, Application US/08877620
Patent No. 5985590
GENERAL INFORMATION:
APPLICANT: Ding, Jeak Ling
TITLE OF INVENTION: The Cloned Factor C cDNA of the
TITLE OF INVENTION: Singapore Horseshoe Crab, Carcinoscopus
TITLE OF INVENTION: rotundicauda and Purification of Factor C Proenzyme
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolaesch & Birch
STREET: 8110 Gatehouse Road, Suite 500 East
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22042
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/877,620
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/596,405
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Murphy, Jr., Gerald M.
REGISTRATION NUMBER: 28,977
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1083 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-877-620-2

Query Match 11.5%; Score 63.5; DB 2; Length 1083;
Best Local Similarity 32.8%; Pred. No. 58;
Matches 19; Conservative 8; Mismatches 22; Indels 9; Gaps 3;

QY 32 MSG--PGCNRAERYSKGCSAIHQKGYDF-STYGTALYNQAGSGVAHTRFGSS 86
DB 378 MSGQPSQVCKVADREVDCSKAV-----DFLDVGEPVRIHCPAGCSLTGTWGTGA 429

RESULT 13
US-07-857-224B-99
Sequence 99, Application US/07857224B
Patent No. 5958784

Qy	24	VNGSAFTYWSG-PCGNRAERYSKCGCSAIHQKGYDFS	YTGQTAA	LYNQACSGVAH--	80
		: : : : : : : : :		:	
Db	1343	VT\$V\$W\$T\$P\$W\$G\$T\$P\$A\$D\$S\$G\$E\$O\$M\$S\$R\$A\$G\$I\$T\$P\$D\$P\$A\$S\$L\$D\$-----	ALARA	VGRRA	GCVTVADID 1398

Mon Jan 13 09:25:42 2003

us-09-882-434a-1.rai

Page 7

OY 81 -TRFGSSARACNP 92
|||:
Db 1399 WERFASAYTATRP 1411

Search completed: January 12, 2003, 09:40:11
Job time : 28 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2003, 09:39:30 ; Search time 20 Seconds

(without alignments)
98.945 Million cell updates/sec

Title: US-09-882-434A-1

Sequence: 1 MASTKLFPSVITVMTLIMAMA.....FGSSARACNPFKWSIFIOC 102

Scoring table: BIOSOM62

Gapop 10.0 , Gapext 0.5

Searched: 118974 seqs, 19401057 residues

Total number of hits satisfying chosen parameters: 118974

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB pep.*
2: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB pep.*
3: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB pep.*
4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB pep.*
5: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB pep.*
6: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB pep.*
7: /cgn2_6/prodata/2/pubpaa/PCUUS_PUBCOMB pep.*
8: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB pep.*
9: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB pep.*
10: /cgn2_6/prodata/2/pubpaa/US09_PUBCOMB pep.*
11: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB pep.*
12: /cgn2_6/prodata/2/pubpaa/US10_PUBCOMB pep.*
13: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB pep.*
14: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	551	100.0	102	10	US-09-882-434A-1 Sequence 1, Appli
2	426	77.3	76	10	US-09-882-434A-15 Sequence 15, Appli
3	426	77.3	76	10	US-09-882-434A-16 Sequence 16, Appli
4	426	77.3	76	10	US-09-882-434A-17 Sequence 17, Appli
5	421	76.4	76	10	US-09-882-434A-19 Sequence 19, Appli
6	419	76.0	76	10	US-09-882-434A-18 Sequence 18, Appli
7	417	75.7	76	10	US-09-882-434A-21 Sequence 21, Appli
8	415	75.3	76	10	US-09-882-434A-20 Sequence 20, Appli
9	75.5	13.7	486	10	US-09-801-368-154 Sequence 154, App
10	70.5	12.8	267	9	US-09-764-868-1164 Sequence 1164, Ap
11	69	12.5	1280	9	US-10-044-671-2 Sequence 2, Appli
12	68	12.3	1280	9	US-10-072-621-7 Sequence 2, Appli
13	68	12.3	1280	10	US-09-866-866A-2 Sequence 2, Appli
14	68	12.3	1280	10	US-09-866-866A-4 Sequence 4, Appli
15	64.5	11.7	147	9	US-10-016-634A-115 Sequence 115, App
16	64	11.6	1272	10	US-09-769-097-2 Sequence 2, Appli
17	64	11.6	1272	10	US-09-769-097-4 Sequence 4, Appli
18	64	11.6	1609	10	US-09-938-275-11 Sequence 11, Appli
19	61	11.1	501	8	US-08-635-967-2 Sequence 2, Appli

20	61	11.1	914	9	US-09-975-143-47	Sequence 47, Appli
21	61	11.1	1786	10	US-09-873-676-113	Sequence 113, App
22	61	11.1	1786	10	US-09-938-275-6	Sequence 6, Appli
23	60.5	11.0	94	9	US-09-950-933A-44	Sequence 44, Appli
24	60.5	11.0	348	10	US-09-791-951-2	Sequence 2, Appli
25	60.5	11.0	787	9	US-09-712-363-178	Sequence 178, App
26	60	10.9	515	10	US-09-895-072-6	Sequence 6, Appli
27	60	10.9	515	10	US-09-986-552-6	Sequence 6, Appli
28	60	10.9	871	10	US-09-893-238-19	Sequence 19, Appli
29	60	10.9	1276	10	US-09-866-866A-8	Sequence 8, Appli
30	60	10.9	1350	10	US-09-893-238-17	Sequence 17, Appli
31	60	10.9	1607	10	US-09-938-275-10	Sequence 10, Appli
32	60	10.9	1786	10	US-09-938-275-7	Sequence 7, Appli
33	60	10.9	2787	10	US-09-893-238-15	Sequence 15, Appli
34	59.5	10.8	861	10	US-09-893-238-12	Sequence 12, Appli
35	59.5	10.8	1404	10	US-09-862-027-24	Sequence 24, Appli
36	59.5	10.8	1743	12	US-10-052-586-451	Sequence 451, App
37	59	10.7	549	10	US-09-893-238-9	Sequence 9, Appli
38	59	10.7	1260	10	US-09-893-238-2	Sequence 2, Appli
39	58.5	10.6	476	12	US-10-078-929-201	Sequence 201, App
40	58.5	10.6	585	10	US-09-841-132-337	Sequence 337, App
41	58.5	10.6	1404	10	US-09-944-849-8	Sequence 8, Appli
42	58.5	10.6	1431	9	US-09-843-930A-2	Sequence 2, Appli
43	58.5	10.6	1752	10	US-09-841-132-180	Sequence 180, App
44	58.5	10.6	5405	9	US-10-023-380-1116	Sequence 1116, Ap
45	58.5	10.6	5405	10	US-09-922-217-1116	Sequence 1116, Ap

ALIGNMENTS

```
RESULT 1
US-09-882-434A-1
; Sequence 1, Application US/09882434A
; Patent No. US20020108144A1
; GENERAL INFORMATION:
; APPLICANT: Manners, John M.
; APPLICANT: Marcus, John Paul
; APPLICANT: Goulter, Kenneth C.
; APPLICANT: Green, Jodie Lynn
; TITLE OF INVENTION: ANTI-MICROBIAL PROTEIN
; FILE REFERENCE: CULIN8.1CPICI
; CURRENT APPLICATION NUMBER: US/09/882,434A
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 09/364395
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 09/117615
; PRIOR FILING DATE: 1998-11-09
; PRIOR APPLICATION NUMBER: PCT/AU97/00052
; PRIOR FILING DATE: 1997-01-31
; PRIOR APPLICATION NUMBER: AU PN 7802
; PRIOR FILING DATE: 1996-01-31
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Macadamia integrifolia
US-09-882-434A-1
Query Match 100.0%; Score 551; DB 10; Length 102;
Best Local Similarity 100.0%; Pred. No. 4.1e-57;
Matches 102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MASTKLFPSVITVMTLIMASEMNGSAFTWSGRCNNRERYSKCCGSAIHQKGYDF 60
DB 1 MASTKLFPSVITVMTLIMASEMNGSAFTWSGRCNNRERYSKCCGSAIHQKGYDF 60
QY 61 SYTGOTALYNOAGCGGVAHTRFSSARACNPFKWSIFIOC 102
DB 61 SYTGOTALYNOAGCGGVAHTRFSSARACNPFKWSIFIOC 102
```

RESULT 2

US-09-882-434A-15
; Sequence 15, Application US/09882434A
; Patent No. US20020108144A1
; GENERAL INFORMATION:
; APPLICANT: Manners, John M.
; APPLICANT: Marcus, John Paul
; APPLICANT: Goulter, Kenneth C.
; APPLICANT: Green, Jodie Lyn
; TITLE OF INVENTION: ANTI-MICROBIAL PROTEIN
; FILE REFERENCE: CULLN18.1CPI1
; CURRENT APPLICATION NUMBER: US/09/882,434A
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 09/364395
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 09/117615
; PRIOR FILING DATE: 1998-11-09
; PRIOR APPLICATION NUMBER: PCT/AU97/00052
; PRIOR FILING DATE: 1997-01-31
; PRIOR APPLICATION NUMBER: AU PN 7802
; PRIOR FILING DATE: 1996-01-31
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Mi28K variant. Variant MiAMP1 protein Mi28K
; OTHER INFORMATION: containing a Lysine at amino acid 28 (used primer
; OTHER INFORMATION: from SEQ ID NO:18 to produce).
US-09-882-434A-15

Query Match 77.3%; Score 426; DB 10; Length 76;
Best Local Similarity 98.7%; Pred. No. 9.3e-43;
Matches 75; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 27 SAFTVWSGPGCNRRAERYSKGCSAIHQKGYDFSVTGQTAAALYNQAGCSGVAHTRFGSS 86
Db 1 SAFTVWSGPGCNRRAERYSKGCSAIHQKGYDFSVTGQTAAALYNQAGCSGVAHTRFGSS 60

Qy 87 ARACNPFQWKSIFIQC 102
Db 61 ARACNPFQWKSIFIQC 76

RESULT 3

US-09-882-434A-16
; Sequence 16, Application US/09882434A
; Patent No. US20020108144A1
; GENERAL INFORMATION:
; APPLICANT: Manners, John M.
; APPLICANT: Marcus, John Paul
; APPLICANT: Goulter, Kenneth C.
; APPLICANT: Green, Jodie Lyn
; TITLE OF INVENTION: ANTI-MICROBIAL PROTEIN
; FILE REFERENCE: CULLN18.1CPI1
; CURRENT APPLICATION NUMBER: US/09/882,434A
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 09/364395
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 09/117615
; PRIOR FILING DATE: 1998-11-09
; PRIOR APPLICATION NUMBER: PCT/AU97/00052
; PRIOR FILING DATE: 1997-01-31
; PRIOR APPLICATION NUMBER: AU PN 7802
; PRIOR FILING DATE: 1996-01-31
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Artificial Sequence

FEATURE:

; OTHER INFORMATION: Mi39K variant. Variant MiAMP1 protein Mi39K
; OTHER INFORMATION: containing a Lysine at amino acid 39 (used primer
; OTHER INFORMATION: from SEQ ID NO:9 to produce).
US-09-882-434A-16

Query Match 77.3%; Score 426; DB 10; Length 76;
Best Local Similarity 98.7%; Pred. No. 9.3e-43;
Matches 75; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 27 SAFTVWSGPGCNRRAERYSKGCSAIHQKGYDFSVTGQTAAALYNQAGCSGVAHTRFGSS 86
Db 1 SAFTVWSGPGCNRRAERYSKGCSAIHQKGYDFSVTGQTAAALYNQAGCSGVAHTRFGSS 60

Qy 87 ARACNPFQWKSIFIQC 102
Db 61 ARACNPFQWKSIFIQC 76

RESULT 4

US-09-882-434A-17
; Sequence 17, Application US/09882434A
; Patent No. US20020108144A1
; GENERAL INFORMATION:
; APPLICANT: Manners, John M.
; APPLICANT: Marcus, John Paul
; APPLICANT: Goulter, Kenneth C.
; APPLICANT: Green, Jodie Lyn
; TITLE OF INVENTION: ANTI-MICROBIAL PROTEIN
; FILE REFERENCE: CULLN18.1CPI1
; CURRENT APPLICATION NUMBER: US/09/882,434A
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 09/364395
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 09/117615
; PRIOR FILING DATE: 1998-11-09
; PRIOR APPLICATION NUMBER: PCT/AU97/00052
; PRIOR FILING DATE: 1997-01-31
; PRIOR APPLICATION NUMBER: AU PN 7802
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Mi46K variant. Variant MiAMP1 protein Mi46K
; OTHER INFORMATION: containing a Lysine at amino acid 46 (used primer
; OTHER INFORMATION: from SEQ ID NO:10 to produce).
US-09-882-434A-17

Query Match 77.3%; Score 426; DB 10; Length 76;
Best Local Similarity 98.7%; Pred. No. 9.3e-43;
Matches 75; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 27 SAFTVWSGPGCNRRAERYSKGCSAIHQKGYDFSVTGQTAAALYNQAGCSGVAHTRFGSS 86
Db 1 SAFTVWSGPGCNRRAERYSKGCSAIHQKGYDFSVTGQTAAALYNQAGCSGVAHTRFGSS 60

Qy 87 ARACNPFQWKSIFIQC 102
Db 61 ARACNPFQWKSIFIQC 76

RESULT 5

US-09-882-434A-19
; Sequence 19, Application US/09882434A
; Patent No. US20020108144A1
; GENERAL INFORMATION:
; APPLICANT: Manners, John M.
; APPLICANT: Marcus, John Paul
; APPLICANT: Goulter, Kenneth C.

APPLICANT: Green, Jodie Lyn
TITLE OF INVENTION: ANTI-MICROBIAL PROTEIN
FILE REFERENCE: CULN18.1CPI1
CURRENT APPLICATION NUMBER: US/09/882,434A
CURRENT FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: 09/364395
PRIOR FILING DATE: 1999-07-30
PRIOR APPLICATION NUMBER: 09/117615
PRIOR FILING DATE: 1998-11-09
PRIOR APPLICATION NUMBER: PCT/AU97/00052
PRIOR FILING DATE: 1997-01-31
PRIOR APPLICATION NUMBER: AU PN 7802
PRIOR FILING DATE: 1996-01-31
NUMBER OF SEQ ID NOS: 21
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 19
LENGTH: 76
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: M154K variant. Variant M154K protein M154V
OTHER INFORMATION: containing a Lysine at amino acid 54 (used primer)
OTHER INFORMATION: from SEQ ID NO:12 to produce).
US-09-882-434A-19

Query Match 76.4%; Score 421; DB 10; Length 76;
Best Local Similarity 98.7%; Pred. No. 3,5e-42;
Matches 75; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 27 SAFTWVGPGCANNRAERYSKGCSAIIHOKGYDPSYTGQTALYNQAGCGSVATRRGSS 86
DB 1 SAFTWVGPGCANNRAERYSKGCSAIIHOKGYDPSYTGQTALYNQAGCGSVATRRGSS 60
QY 87 ARACNPFGMKSIPIQC 102
DB 61 ARACNPFGMKSIPIQC 76

RESULT 6
US-09-882-434A-18
Sequence 18, Application US/09882434A
Patent No. US20020108144A1
GENERAL INFORMATION:
APPLICANT: Manners, John M.
APPLICANT: Marcus, John Paul
APPLICANT: Goulter, Kenneth C.
APPLICANT: Green, Jodie Lyn
TITLE OF INVENTION: ANTI-MICROBIAL PROTEIN
FILE REFERENCE: CULN18.1CPI1
CURRENT APPLICATION NUMBER: US/09/882,434A
CURRENT FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: 09/364395
PRIOR FILING DATE: 1999-07-30
PRIOR APPLICATION NUMBER: 09/117615
PRIOR FILING DATE: 1998-11-09
PRIOR APPLICATION NUMBER: PCT/AU97/00052
PRIOR FILING DATE: 1997-01-31
PRIOR APPLICATION NUMBER: AU PN 7802
PRIOR FILING DATE: 1996-01-31
NUMBER OF SEQ ID NOS: 21
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 18
LENGTH: 76
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: M154V variant. Variant M154K protein M154V
OTHER INFORMATION: containing a Valine at amino acid 54 (used primer)
OTHER INFORMATION: from SEQ ID NO:11 to produce).
US-09-882-434A-18

Query Match 76.0%; Score 419; DB 10; Length 76;
Best Local Similarity 98.7%; Pred. No. 6e-42;

Matches 75; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 27 SAFTWVGPGCANNRAERYSKGCSAIIHOKGYDPSYTGQTALYNQAGCGSVATRRGSS 86
DB 1 SAFTWVGPGCANNRAERYSKGCSAIIHOKGYDPSYTGQTALYNQAGCGSVATRRGSS 60
QY 87 ARACNPFGMKSIPIQC 102
DB 61 ARACNPFGMKSIPIQC 76

RESULT 7
US-09-882-434A-21
Sequence 21, Application US/09882434A
Patent No. US20020108144A1
GENERAL INFORMATION:
APPLICANT: Manners, John M.
APPLICANT: Marcus, John Paul
APPLICANT: Goulter, Kenneth C.
APPLICANT: Green, Jodie Lyn
TITLE OF INVENTION: ANTI-MICROBIAL PROTEIN
FILE REFERENCE: CULN18.1CPI1
CURRENT APPLICATION NUMBER: US/09/882,434A
CURRENT FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: 09/364395
PRIOR FILING DATE: 1999-07-30
PRIOR APPLICATION NUMBER: 09/117615
PRIOR FILING DATE: 1998-11-09
PRIOR APPLICATION NUMBER: PCT/AU97/00052
PRIOR FILING DATE: 1997-01-31
PRIOR APPLICATION NUMBER: AU PN 7802
PRIOR FILING DATE: 1996-01-31
NUMBER OF SEQ ID NOS: 21
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 21
LENGTH: 76
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: M146K/54K variant. Variant M146K protein M146K/54K containing a Lysine at amino acid 46 and
OTHER INFORMATION: a Lysine at amino acid 54.
US-09-882-434A-21

Query Match 75.7%; Score 417; DB 10; Length 76;
Best Local Similarity 97.4%; Pred. No. 1e-41;
Matches 74; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 27 SAFTWVGPGCANNRAERYSKGCSAIIHOKGYDPSYTGQTALYNQAGCGSVATRRGSS 86
DB 1 SAFTWVGPGCANNRAERYSKGCSAIIHOKGYDPSYTGQTALYNQAGCGSVATRRGSS 60
QY 87 ARACNPFGMKSIPIQC 102
DB 61 ARACNPFGMKSIPIQC 76

RESULT 8
US-09-882-434A-20
Sequence 20, Application US/09882434A
Patent No. US20020108144A1
GENERAL INFORMATION:
APPLICANT: Manners, John M.
APPLICANT: Marcus, John Paul
APPLICANT: Goulter, Kenneth C.
APPLICANT: Green, Jodie Lyn
TITLE OF INVENTION: ANTI-MICROBIAL PROTEIN
FILE REFERENCE: CULN18.1CPI1
CURRENT APPLICATION NUMBER: US/09/882,434A
CURRENT FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: 09/364395
PRIOR FILING DATE: 1999-07-30
PRIOR APPLICATION NUMBER: 09/117615

; PRIOR FILING DATE: 1998-11-09
; PRIOR APPLICATION NUMBER: PCT/AU97/00052
; PRIOR FILING DATE: 1997-01-31
; PRIOR APPLICATION NUMBER: AU PN 7802
; PRIOR FILING DATE: 1996-01-31
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: M146K/54V variant. Variant M146K protein
; OTHER INFORMATION: M146K/54V containing a Lysine at amino acid 46 and
; OTHER INFORMATION: a Valine at amino acid 54.
US-09-882-434A-20

Query Match 75.3%; Score 415; DB 10; Length 76;
Best Local Similarity 97.4%; Pred. No. 1.8e-41;
Matches 74; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 27 SAFTVWSGPGCNRAERYSKGCSAIIHQGGYDFSYTGQTAAALYNQAGCGVAHTRFGSS 86
DDB 1 SAFTVWSGPGCNRAERYSKGCSAIIHQGGYDFSYTGQTAAALYNQAGCGVAHTRFGSS 60
QY 87 ARACNPFQWKSIFITQC 102
DDB 61 ARACNPFQWKSIFITQC 76

RESULT 9
US-09-801-368-154
; Sequence 154, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 154
; LENGTH: 486
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-154

Query Match 13.7%; Score 75.5; DB 10; Length 486;
Best Local Similarity 30.2%; Pred. No. 0.35;
Matches 29; Conservative 9; Mismatches 27; Indels 31; Gaps 6;
QY 8 FSVLTVMV---LIAMSEWVNGSAFTVWSGPGCNRAERYSKGCSAIIHQGGYDFSYTG 64
DDB 369 FGINTVQERKLIRRLSELIGA-----RAARLSVCGIAAICQKRGYK---TG 412
QY 65 QTAALYNQAGCGVAHTRF-GSSARACNP-----FGW 95

Db 413 HIAA-----DGSVYNRYPGFKKAANALKDIYCW 441

RESULT 10
US-09-764-868-1164
; Sequence 1164, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1164
; LENGTH: 267
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (115)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-868-1164

Query Match 12.8%; Score 70.5; DB 9; Length 267;
Best Local Similarity 27.0%; Pred. No. 0.66;
Matches 20; Conservative 13; Mismatches 36; Indels 5; Gaps 2;
QY 28 AFTVWSGPGCNRAERYSKGCSAIIHQGGYDFSYTGQTAAALYNQAGCGVAHTRFGSSA 87
DDB 146 SFEFWDSPFNDSRGRGH--GLRGAFSAGFGFEPAFMEAFSFFNMLGCGSGSHITFSSTS 203
QY 88 ---RACNPFQWKS 98
DDB 204 FCGSSGSGSGFKSV 217

RESULT 11
US-10-044-671-2
; Sequence 2, Application US/10044671
; Patent No. US20020177147A1
; GENERAL INFORMATION:
; APPLICANT: Washington State University Research Foundation
; APPLICANT: Mealey, Katrina
; APPLICANT: Bentjen, Steven
; TITLE OF INVENTION: MDR1 VARIANTS AND METHODS FOR THEIR USE
; FILE REFERENCE: 4630-61733
; CURRENT APPLICATION NUMBER: US/10/044,671
; CURRENT FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/261,578
; PRIOR FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: US 60/314,829
; PRIOR FILING DATE: 2001-08-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1280
; TYPE: PRT
; ORGANISM: Canis familiaris
US-10-044-671-2

Query Match 12.5%; Score 69; DB 9; Length 1280;
Best Local Similarity 25.5%; Pred. No. 6;
Matches 25; Conservative 12; Mismatches 33; Indels 28; Gaps 3;
QY 6 LFPSVITVWMLIAMSEMV-----NGSAFTVWSGPGCNRAERYSKGCSAIIHQGGYD 59
DDB 335 VFPSLIGAFSGQASPSIEAFANARGAAYEIFKIIDNKPSIDSYSGSKGHPDNKIGNLE 394
QY 60 -----FSY-----TGQTAAALYNQAGC 75

Db ||||| ||
 65 GSSAWPDSTCN 75

Search completed: January 12, 2003, 09:46:46
Job time : 22 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2003, 09:36:00 ; Search time 43 Seconds
(without alignments)

228.040 Million cell updates/sec

Title: US-09-882-434a-1

Sequence: 1 MASTKLFPSVITVMTLMAA.....FGSSARACNPFQWKSIFIOC 102

Scoring table: BIOSUM62
Gapop 10.0, Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	75.5	13.7	315	2 T35804	probable aldehyde
2	75.5	13.7	486	1 KIBYHB	hexokinase (EC 2.7
3	69.5	12.6	532	2 AB3552	alcohol dehydrogen
4	69.5	12.5	1281	2 T48123	p-glycoprotein iso
5	68.5	12.4	1557	2 T28811	hypothetical prote
6	68.5	12.3	1280	1 DVHUI	multidrug resistanc
7	67.5	12.3	184	2 T16044	hypothetical prote
8	66.5	12.1	81	2 T02667	proteinase inhibit
9	66.5	12.1	151	2 JCG040	fibrin protein ag
10	66.5	12.1	157	2 S31078	seed allergen RA5
11	66.5	12.1	152	2 S21157	seed allergen RA17
12	66.5	12.1	251	2 T34656	hypothetical prote
13	66.5	12.1	461	2 AE0311	MADH2 dehydrogenas
14	66.5	12.0	1276	1 DVMS2	multidrug resistanc
15	65.5	11.9	151	2 AH0635	nucleation compone
16	65.5	11.9	233	2 T23296	hypothetical prote
17	65.5	11.9	244	2 T04212	osmotic precursor
18	65.5	11.8	165	2 S31080	seed allergen RA14
19	64.5	11.7	485	1 KIBYHA	hexokinase (EC 2.7
20	64.5	11.6	320	2 S00935	rx protein - Stap
21	64.5	11.6	1278	2 S41646	p-glycoprotein - r
22	64.5	11.6	1551	2 AB2410	WD-repeat protein
23	64.5	11.6	1609	1 KIBYHB	laminin gamma-1 ch
24	63.5	11.5	3016	2 S77300	hypothetical prote
25	63.5	11.5	166	2 S31082	seed allergen RA62
26	63.5	11.5	503	2 S31940	starch-degrading e
27	63.5	11.5	503	2 C97262	N-terminal domain
28	63.5	11.5	828	1 D39142	outer membrane ush
29	63.5	11.5	1019	2 A38738	coagulation factor

ALIGNMENTS

30	63	11.4	157	2 T02664	allergen - rice
31	63	11.4	159	2 S26238	pathogenesis-relat
32	63	11.4	339	2 A47611	env polypeptid -
33	63	11.4	1276	1 DVHVIC	multidrug resistanc
34	63	11.4	1279	1 DVHVIC	multidrug resistanc
35	62.5	11.3	125	2 T16247	hypothetical prote
36	62.5	11.3	166	2 S59922	allergen RA14B pre
37	62.5	11.3	168	2 T07146	pathogenesis-relat
38	62.5	11.3	235	1 OTTC2	chaumatin II precu
39	62.5	11.3	779	2 H71301	probable membrane
40	62	11.3	159	1 VCT014	pathogenesis-relat
41	62	11.3	159	2 S00900	heat shock 18k pro
42	62	11.3	234	2 T21495	hypothetical prote
43	62	11.3	340	2 T13781	MADH2 dehydrogenas
44	62	11.3	461	2 C86935	probable carboxype
45	62	11.3	553	2 T52362	hypothetical prote

RESULT 1

T35804
probable aldehyde dehydrogenase - Streptomyces coelicolor (fragment)

C/Species: Streptomyces coelicolor

C/Date: 05-Nov-1999 #sequence revision 05-Nov-1999 #text_change 05-Nov-1999

C/Accession: T35804

R/Murphy, L.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, August 1998

A/Reference number: Z21556

A/Accession: T35804

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-315 <MUR>

A/Cross-references: EMBL:AL031225; PIDN:CAA20223.1; GSPDB:GN00070; SCODEB:SC8B7.12c

C/Experimental source: strain A3(12)

C/Genetics:

A/Gene: SCODEB:SC8B7.12c

Query Match	13.7%	Score 75.5;	DB 2,	Length 315;
Best Local Similarity	25.9%	Pred. No. 2.4;		
Matches	30;	Conservative 20;	Mismatches 47;	Indels 19;
			Gaps	5
QY	1	MASTKLFPSVITVMTL-----TAMASEMNGSAFTVWGPGCN-NRARRYSKCGSATH	53	
	::	:	:
DB	199	VQSEIFGPVIVLPDPTDEGIRLNDPRYGLAASMSRDVRRARRRIRAGCWAIN	258	
QY	54	Q-----KGGVDFSYGTGTALVYQACSGVAHTRFGSSARACNPFQW-KSIF	99	
	::	:	:
DB	259	DHPIIEMPHGGVYKASGFGKMSAISFEYTVQKVMFNTVAIAK--DMHRTVF	312	

RESULT 2

KIBYHB
hexokinase (EC 2.7.1.1) B - yeast (Saccharomyces cerevisiae)

N/Alternate names: HEX1 protein; hexokinase II; hexokinase PII; protein G0556; protein

C/Species: Saccharomyces cerevisiae

C/Date: 28-Dec-1987 #sequence revision 12-Apr-1996 #text_change 16-Jun-2000

C/Accession: S61608; B23523; S22430; A23958; S05731; S33656; S64279; A53632; S28555

R/Colisac E; Maillet E; Robineau S; Netter P.

submitted to the EMBL Data Library, December 1995

A/Reference number: S61598

A/Accession: S61608

A/Molecule type: DNA

A/Residues: 1-486 <COI>

A/Cross-references: EMBL:X94357; NID:G1150575; PIDN:CAA64134.1; PID:G1150586

R/Stachel C; Stachel J; Swan J; Botstein D; Konigsberg W.

Nucleic Acids Res 14, 945-963, 1986

A/Title: Identification, cloning and sequence determination of the genes specifying hexa

A/Reference number: A93649; MUID:6120382; PMID:3003701

A/Accession: B23523

A/Molecule type: DNA

A/Residues: 1-28, 'I', '30-60, 'V', '62-196, 'S', '198-486 <STA>

A;Cross-references: EMBL:X03483; NID:g3792; PIDN:CAA27203.1; PID:g3793
A;Note: the authors translated the codon GTT for residue 61 as Gly
A;Accession: S22430
A;Molecule type: protein
A;Residues: 2-7; I: 30-46; 78-111; 114-119; 166-173; 177-194; 228-236; 250-281; 283-300; 330-335
R;Prolich, K.; Entian, K.; Mecke, D.
Gene 36, 105-111, 1985
A;Title: The primary structure of the yeast hexokinase PII gene (HXK2) which is responsible for the regulation of glucose uptake in *Saccharomyces cerevisiae*
A;Reference number: A23958; MUID:86056943; PMID:3905511
A;Accession: A23958
A;Molecule type: DNA
A;Residues: 1-32; N: 34-60; V: 62-420; ST: 423-443; PH: 446-452; V: 454-461; P: 463-486
A;Cross-references: EMBL:M11181
A;Note: the authors translated the codon GTT for residue 61 as Gly
R;Schmidt, J.J.; Colowick, S.P.
Arch. Biochem. Biophys. 158, 471-477, 1973
A;Title: Identification of a peptide sequence involved in association of subunits of yeast hexokinase
A;Reference number: S05731; MUID:74114889; PMID:4592981
A;Accession: S05731
A;Molecule type: protein
A;Residues: 2-12 <SCH>
R;Breitwieser, W.; Price, C.; Schuster, T.
Yeast 9, 551-556, 1993
A;Title: Identification of a gene encoding a novel zinc finger protein in *Saccharomyces cerevisiae*
A;Reference number: S33654; MUID:93311123; PMID:8322518
A;Accession: S33654
A;Molecule type: DNA
A;Residues: 1-247 <BRE>
A;Cross-references: EMBL:X67787; NID:g3707; PIDN:CAA48003.1; PID:g3710
R;Coissac, E.; Maillier, E.; Netter, P.
submitted to the Protein Sequence Database, May 1996
A;Reference number: S64271
A;Accession: S64271
A;Molecule type: protein
A;Residues: 1-486 <CON>
A;Cross-references: EMBL:272775; NID:gl322930; PIDN:CAA96973.1; PID:gl322931; GSPDB:GN00191
A;Experimental source: strain S288C
R;Kriegel, T.M.; Rush, J.; Vojtek, A.B.; Clifton, D.; Fraenkel, D.G.
Biochemistry 33, 148-152, 1994
A;Title: In vivo phosphorylation site of hexokinase 2 in *Saccharomyces cerevisiae*
A;Reference number: A53632; MUID:94114477; PMID:8286332
A;Accession: A53632
A;Status: preliminary
A;Molecule type: protein
A;Residues: 13-21 <KRI>
C;Genetics:
A;Gene: SGD.HXK2; HEX1; SC12; HXB; MIPS:YGL253W
A;Cross-references: SGD:S0003222; MIPS:YGL253W
A;Map position: 7L
C;Superfamily: hexokinase; hexokinase homology
C;Keywords: allosteric regulation; ATP; glycolysis; phosphoprotein; phosphotransferase
F:2-486/Product: hexokinase B #status experimental <MAT>
F:36-470/Domain: hexokinase homology <HXK>
F:15/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match 13.7%; Score 75.5; DB 1; Length 486;
Best Local Similarity 30.2%; Pred. No. 3.5;
Matches 29; Conservative 9; Mismatches 27; Indels 31; Gaps 6;
QY 8 FSVITVM---LIAMSEWNGSAFTVWSGPGCNRAERYSKGCSAIHQKGYDFS YTG 64
DB 369 FGINTVQERKLRLRLSLIGA-----RAARLSVCGIATCQKRGYK---TG 412
QY 65 QTAALYNQAGCGVAHTRF-GSSARACNP---FGW 95
DB 413 HIAA-----DGSVYNYRPFKEKAALAKDIYGW 441

RESULT 3
AB3552
alcohol dehydrogenase (acceptor) (EC 1.1.1.99.-) [imported] - *Brucella melitensis* (strain
C;Species: *Brucella melitensis*
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002

C;Accession: AB3552
R;DelVecchio, V.G.; Kaputral, V.; Redkar, R.J.; Patra, G.; Mijer, C.; Los, T.; Ivanova, J.; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letessier, P.; Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A;Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*
A;Reference number: AD3252; PMID:11756688
A;Accession: AB3552
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-532 <KUR>
A;Cross-references: GB:AE008918; PIDN:AAL53581.1; PID:gl7984492; GSPDB:GN00191
A;Experimental source: strain 16M
C;Genetics:
A;Gene: BMEI10339
A;Map position: II
C;Superfamily: alcohol oxidase
C;Keywords: oxidoreductase

Query Match 12.6%; Score 69.5; DB 2; Length 532;
Best Local Similarity 27.2%; Pred. No. 16;
Matches 22; Conservative 8; Mismatches 16; Indels 35; Gaps 4;
QY 29 FTVWSGPGCNRAERYSK-----CGCSAIH-----OKGYDFS YTGTAALYNQAGCSG 77
DB 64 FTTEAEPLNGRSLNTPRGKVLGGCSINGMIYMRGQACDYD-----LWRQAGCDG 114
QY 78 VAHTRFGSSARACNPFQWKS I 98
DB 115 -----NGWDDV 120

RESULT 4
I48123
p-glycoprotein isoform III - Chinese hamster
C;Species: *Cricetulus griseus* (Chinese hamster)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 02-Feb-2001
C;Accession: I48123
R;Endicott, J.A.; Sarangi, P.; Ling, V.
DNA Seq. 2, 89-101, 1991
A;Title: Complete cDNA sequences encoding the Chinese hamster P-glycoprotein gene family
A;Reference number: I48121; MUID:92135896; PMID:1685679
A;Accession: I48123
A;Status: preliminary; translated from GB/EMBL/DBS
A;Molecule type: mRNA
A;Residues: 1-1281 <RES>
A;Cross-references: GB:M60042; NID:gl91168; PIDN:AAA68885.1; PID:gl91169
C;Superfamily: multidrug resistance protein; ATP-binding cassette homology
C;Keywords: ATP; glycoprotein; nucleotide binding; P-loop
F:412-606/Domain: ATP-binding cassette homology <ABCI>
F:429-436/Region: nucleotide-binding motif A (P-loop)
F:1054-1250/Domain: ATP-binding cassette homology <ABC2>
F:1071-1078/Region: nucleotide-binding motif A (P-loop)

Query Match 12.5%; Score 69; DB 2; Length 1281;
Best Local Similarity 21.4%; Pred. No. 41;
Matches 21; Conservative 17; Mismatches 32; Indels 28; Gaps 2;
QY 6 LFPSVITVMMLIAMSEWVN-----GSAFTVWSGPGCNRAERYSKGCSAIHQKGYD 59
DB 336 VFPSILIGAFVGOAPCIDAFANARGAAYVIFDIIDNNPKIDSFSERGHKPDISKGNLD 395
QY 60 FS-----YTGTAALYNQAGC 75
DB 396 FSDVHFSYPSRANIKLGLNLKVKSGQTVALVGNSGC 433

RESULT 5
T28811
hypothetical protein C54D1.5 - *Caenorhabditis elegans*
C;Species: *Caenorhabditis elegans*
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C;Accession: T28811
R;Minx, M.

submitted to the EMBL Data Library, January 1996
A:Description: The sequence of C. elegans cosmid C54D1.
A:Reference number: Z20527
A:Accession: T28811
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1557 <MIN>
A:Cross-references: EMBL:U46673; PIDN:AA048152.1; GSPDB:GNO0028; CESP:C54D1.5
A:Experimental source: strain Bristol N2; clone C54D1
A:Gene: CESP:C54D1.5
A:Map position: X
A:Introns: 84/3; 127/1; 166/2; 271/2; 331/1; 392/3; 433/2; 585/2; 1089/1; 1530/3
C:Superfamily: laminin beta-1 chain; laminin-type EGF-like homology

Query Match 12.4%; Score 68.5; DB 2; Length 1557;
Best Local Similarity 28.8%; Pred. No. 55;
Matches 19; Conservative 5; Mismatches 33; Indels 9; Gaps 2;

QY 29 FTVMSGPCNNNAERYSKGCSAHHQKGYDPSYTGCTALYNQAG--CSGVATRRFGSS 86
DB 920 YNTSGDLC-----QECNCDPLSGEVTCDVNTGCCCKKPGVVGQRCDRCADYHFGFS 972

QY 87 ARACNP 92
DB 973 ANGCOP 978

RESULT 6
DVTU1
multidrug resistance protein 1 - human
N:Alternate names: P-glycoprotein 1
C:Species: Homo sapiens (man)
C>Date: 31-Dec-1990 #sequence_revision 18-Aug-1995 #text_change 19-Jan-2001
R:Accession: A34914; PS0162; S15509; S43838; I52238; I55204
R:Chen, C.; Clark, D.; Ueda, K.; Pastan, I.; Gottesman, M.M.; Roninson, I.B.
J. Biol. Chem. 265, 506-514, 1990
A:Title: Genomic organization of the human multidrug resistance (MDR1) gene and origin
A:Reference number: A34914; MUID:90094448; PMID:1967175

A:Molecule type: DNA
A:Residues: 1-1280 <CHE>
A:Cross-references: GB:M29447; GB:J05168; NID:G187496; PIDN:AAA59576.1; PID:G386862
R:Kloka, N.; Yamano, Y.; Komano, T.; Ueda, K.
submitted to JEPD, April 1991
A:Reference number: PS0162
A:Accession: PS0162

A:Molecule type: DNA
A:Residues: 1-22 <KIO>
R:Kloka, N.; Yamano, Y.; Komano, T.; Ueda, K.
submitted to the EMBL Data Library, April 1991
A:Description: Transcriptional regulation of multidrug resistance gene (MDR1) expression
A:Reference number: S15500
A:Accession: S15500

A:Molecule type: DNA
A:Residues: 1-22, 'R' <K12>
A:Cross-references: EMBL:X58723; NID:G34522; PIDN:CAA1558.1; PID:G34523
R:Chen, C.; Chin, J.B.; Ueda, K.; Clark, D.P.; Pastan, I.; Gottesman, M.M.; Roninson, I.
Cell 47, 381-389, 1986
A:Title: Internal duplication and homology with bacterial transport proteins in the mdr1
A:Reference number: A25059; MUID:87028230; PMID:2876781
A:Accession: A25059

A:Molecule type: mRNA
A:Residues: 1-184, 'V' 186-1280 <CH2>
A:Cross-references: GB:M14758; NID:G187468; PIDN:AAA59575.1; PID:G307180
R:Chambers, T.C.; Pohl, J.; Glass, D.B.; Kuo, J.F.
Biochem. J. 299, 309-315, 1994
A:Title: Phosphorylation by protein kinase C and cyclic AMP-dependent protein kinase of
A:Reference number: S43838; MUID:94220047; PMID:7909431
A:Accession: S43838

A:Molecule type: protein
A:Residues: 656-689 <CHA>
R:Geiseler, V.; Weger, S.; Probst, H.

Biochem. Biophys. Res. Commun. 169, 796-802, 1990
A:Title: mdr1/P-glycoprotein gene segments analyzed from various human leukemic cell lines
A:Reference number: I52238; MUID:90250529; PMID:1572623
A:Accession: I52238
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 178-215 <RES>
A:Cross-references: GB:M37724; NID:G183537; PIDN:AAA8047.1; PID:G553314
A:Accession: I55204

A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 800-856 <RE2>
A:Cross-references: GB:M37725; NID:G183538; PIDN:AAA8048.1; PID:G553315
C:Comment: This is an integral membrane protein overproduced in multidrug-resistant cell
structurally and functionally unrelated lipophilic anticancer drugs.
C:Genetics:
A:Gene: GDB:PCY1, MDR1
A:Cross-references: GDB:120712; OMIM:171050
A:Map position: 7q21-7q21
C:Superfamily: multidrug resistance protein; ATP-binding cassette homology
C:Keywords: ATP; duplication; glycoprotein; nucleotide binding; P-loop; phosphoprotein;
F:1-638, 653-1280/Region: duplication
F:49-350/Domain: hydrophobic <HB1>
F:351-637/Domain: hydrophilic <HL1>
F:410-634/Region: ATP-binding cassette homology <ABC1>
F:427-434/Region: nucleotide-binding motif A (P-loop)
F:551-555/Region: nucleotide-binding motif B
F:638-708/Domain: linker <LIN>
F:709-993/Domain: hydrophobic <HB2>
F:994-1280/Domain: hydrophilic <HL2>
F:1053-1249/Domain: ATP-binding cassette homology <ABC2>
F:1070-1077/Region: nucleotide-binding motif A (P-loop)
F:1196-1200/Region: nucleotide-binding motif B
F:91, 94, 99/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:433/Binding site: ATP (Lys) #status predicted
F:661, 667, 671/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status ext
F:667, 671, 683/Binding site: phosphate (Ser) (covalent) (by cAMP-dependent kinase) #status
F:1076/Binding site: ATP (Lys) #status predicted

Query Match 12.3%; Score 68; DB 1; Length 1280;
Best Local Similarity 24.5%; Pred. No. 51;
Matches 24; Conservative 13; Mismatches 33; Indels 28; Gaps 3;

QY 6 LFPSVTVMMLTAMASENV-----NSAFTWSGPCNNRAERYSKGCSAHHQKGYD 59
DB 334 VFPSVLIGAFSVGQASPSLEAFAMRGAAVEFKLIIDNKPSIDISKSGHKPKDNIKGLDE 393

QY 60 -----FSY-----TGCTAALYNQAG 75
DB 394 FRNHFSPYSRKEVKILKGLINKVQSGTVALVNSGC 431

RESULT 7
T16044
hypothetical protein F12A10.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T16044
R:Geiseler, C.
submitted to the EMBL Data Library, June 1995
A:Description: The sequence of C. elegans cosmid F12A10.
A:Reference number: Z18451
A:Accession: T16044

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-184 <GRT>
A:Cross-references: EMBL:U28731; NID:G861241; PID:G861242; PIDN:AAA68294.1; CESP:F12A10
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:F12A10.1
A:Introns: 25/2; 66/1, 114/2

Query Match 12.3%; Score 67.5; DB 2; Length 184;

Best Local Similarity 24.1%; Pred. No. 9.6;
Matches 26; Conservative 11; Mismatches 34; Indels 37; Gaps 4;
QY 1 MASTKLFFSVITVWMLIAMASEMVN-----CSAFTVWSG-----PG 36
Db 78 MHSTTLFFVVFGLVAVSLAFDDLNHPKNLQWRSSGSRVLRWGGGGYPGGYGGGYG 137
QY 37 CNNRAERYSKGCSATHOKGDFSYTGQTAALYNQAGCGVAHTRFG 84
Db 138 -----GYSGY-----GGYFGYGGGGYSGSSWGSYSSRSRG 172
RESULT 8
T02667
protease inhibitor - rice
C:Species: Oryza sativa (rice)
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 09-Jun-2000
C:Accession: T02667
R:Yun, C.H.; Lee, J.H.; Park, J.H.; Lee, G.R.
Submitted to the EMBL Data Library, January 1998
A:Description: Molecular characterization of rice proteinase inhibitor gene.
A:Reference number: Z14694
A:Accession: T02667
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-81 <YUN>
A:Cross-references: EMBL:AF044059; NID:g2829211; PIDN:AAC00503.1; PID:g2829212
A:Experimental source: strain IR36
C:Genetics:
A:Gene: RGP19
A:Introns: 21/1
C:Superfamily: gamma-thionin
Query Match 12.1%; Score 66.5; DB 2; Length 81;
Best Local Similarity 23.2%; Pred. No. 5.7;
Matches 19; Conservative 16; Mismatches 28; Indels 19; Gaps 3;
QY 2 ASTKLFFSVITVWMLIAMASEMVNGSAFTVWSGPGCNNAERYS-----KGCSSAIHQK 56
Db 3 ASRKVFSAMLLMWLLLAATGEM--GGPVWAEARTCESQSHRFKGPCKARKANCASVCNTE 60
QY 57 GYDFSYTGQTAALYNQAGCGSV 78
Db 61 GPPDGY-----CHGV 70
RESULT 9
JC6040
fimbriin protein agfB precursor - Salmonella enteritidis
C:Species: Salmonella enteritidis
C:Date: 31-Dec-1996 #sequence_revision 31-Dec-1996 #text_change 08-Oct-1999
C:Accession: JC6040
R:Collinson, S.K.; Clouthier, S.C.; Doran, J.L.; Banser, P.A.; Kay, W.W.
J. Bacteriol. 178, 662-667, 1996
A:Title: Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae.
A:Reference number: JC6039; MUID:96146512; PMID:8550497
A:Accession: JC6040
A:Molecule type: DNA
A:Residues: 1-151 <COL>
A:Cross-references: GB:U43280; NID:gl184712; PIDN:AAC43598.1; PID:gl184713
A:Experimental source: strain 276755-3b
C:Genetics:
A:Gene: agfB
A:Function:
A:Description: minor component of thin aggregative fimbriae
A:Note: fimbriae bind to fibronectin, plasminogen, tissue plasminogen activator
C:Keywords: fimbria
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-151/Product: fimbriin protein agfB #status predicted <MAT>
Query Match 12.1%; Score 66.5; DB 2; Length 151;
Best Local Similarity 31.0%; Pred. No. 10;
Matches 18; Conservative 11; Mismatches 16; Indels 13; Gaps 2;

QY 30 TVWSGPGCNNAERYSKGCSATHOKGDFSYTGQTAALYNQAGCGVAHTRFGSSA 87
Db 69 SVISOEGGNRAK-----VDQAGNYNFAVTEQT-----GNANDASISQSAYGNSA 113
RESULT 10
S31078
seed allergen RA5 - rice
C:Species: Oryza sativa (rice)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 20-Jun-2000
C:Accession: S31078
R:Adachi, T.; Izumi, H.; Yamada, T.; Tanaka, K.; Takeuchi, S.; Nakamura, R.; Mateuda, T.
Plant Mol. Biol. 21, 239-248, 1993
A:Title: Gene structure and expression of rice seed allergenic proteins belonging to the
A:Reference number: S31078; MUID:93144699; PMID:7678765
A:Accession: S31078
A:Molecule type: mRNA
A:Residues: 1-157 <ADA>
A:Cross-references: EMBL:D11430; NID:g218196; PIDN:BAAO1996.1; PID:g218197
C:Superfamily: wheat alpha-amylase inhibitor
C:Keywords: seed
Query Match 12.1%; Score 66.5; DB 2; Length 157;
Best Local Similarity 22.1%; Pred. No. 11;
Matches 33; Conservative 12; Mismatches 35; Indels 69; Gaps 7;
QY 1 MASTKLFFSV--ITVWMLIAMASEMVN-----GSAFTVWSGPGCNNAERAE 42
Db 1 MASNKVVFVSVLLLVAVSVLAATATMAEYHQDQVYVTRARCPQGMGYPMYSLPRCALVK 60
QY 43 RY-----SKGCSAI-HQKGYDFSYTGQTAALYNQAGCGS 77
Db 61 RQCRGSAAAEQVRDCCRQLAAVDDSMCRCEATSHMLGG-----IYRELGAPD 109
QY 78 VAH-----TRFGSSARA-----CN 91
Db 110 VGHPWSEVPRGCRGDLERAAASLPAFCN 138
RESULT 11
S21157
seed allergen RA17 - rice
C:Species: Oryza sativa (rice)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 20-Jun-2000
C:Accession: S21157; JC4887; S31079
R:Izumi, H.; Adachi, T.; Fujii, N.; Matsuda, T.; Nakamura, R.; Tanaka, K.; Uriasu, A.; Ku:
FEBS Lett. 302, 213-216, 1992
A:Title: Nucleotide sequence of a cDNA clone encoding a major allergenic protein in rice
A:Reference number: S21157; MUID:92289999; PMID:1376283
A:Accession: S21157
A:Molecule type: mRNA
A:Residues: 1-162 <IZU>
A:Cross-references: EMBL:D11431; NID:g218194; PIDN:BAAO1997.1; PID:g218195
R:Nakamura, R.; Mateuda, T.
Biosci. Biotechnol. Biochem. 60, 1215-1221, 1996
A:Title: Rice allergenic protein and molecular-genetic approach for hypoallergenic rice.
A:Reference number: JC4887; MUID:97141195; PMID:8987539
A:Contents: seed
A:Accession: JC4887
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-162 <NAK>
A:Cross-references: EMBL:X66257; NID:g311892; PIDN:CAA46983.1; PID:g311893
C:Comment: This protein shows a sequence similarity to alpha-amylase/trypsin inhibitor f.
C:Superfamily: wheat alpha-amylase inhibitor
Query Match 12.1%; Score 66.5; DB 2; Length 162;
Best Local Similarity 21.5%; Pred. No. 11;
Matches 31; Conservative 17; Mismatches 37; Indels 59; Gaps 7;
QY 1 MASTKLFFSVITVWMLIAMASEMVN-----GSAFTVWSGPGCNNAERY 44

1. Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.
 2. Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* sero-
 3. Reference number: AB0502; PMID:11677608

us-09-882-434a-1.rpr

Mon Jan 13 09:25:46 2003

A;Accession: AH0635
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-151 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD08267.1; PID:g16502314; GSPDB:GN00176
C;Genetics:
A;Gene: STY1180

Query Match 11.9%; Score 65.5; DB 2; Length 151;
Best Local Similarity 31.0%; Pred. No. 13;
Matches 18; Conservative 11; Mismatches 16; Indels 13; Gaps 2;
Qy 30 TVWSGPCNNRAERYSKGCSAIHQGGYDFSYTGQTAAALYNQAGCSGVAHTRFGSSA 87
Db 69 SVISQEGENNRK-----VDQAGNYNFAYIEQT-----GNANDASISQSAVGNSA 113

Search completed: January 12, 2003, 09:39:24
Job time : 45 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2003, 08:42:05 ; Search time 25 Seconds

(without alignments)
169.223 Million cell updates/sec

Title: US-09-882-434a-1

Perfect score: 551
Sequence: 1 MASTKLFPSVITVMMILMAA.....FGSSARACNPFMGKSIPIQC 102

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	551	100.0	102	1	AMP1 MACIN
2	75.5	13.7	485	1	HXKB YEAST
3	71.5	13.0	403	1	PGK CHLNU
4	69	12.5	1281	1	MDR3 CRIGR
5	68.5	12.4	1557	1	LMU1 CAEEL
6	68	12.3	1280	1	MDR1 HUMAN
7	67.5	12.3	902	1	VEF GVHA
8	66.5	12.1	157	1	CSGB SALTY
9	66.5	12.1	157	1	RA05 ORYSA
10	66.5	12.1	162	1	RA17 ORYSA
11	66	12.0	1276	1	MDR2 MOUSE
12	65.5	11.9	151	1	CSGB SALTY
13	65.5	11.9	244	1	OSL3 ARATH
14	65	11.8	165	1	RA14 ORYSA
15	64.5	11.7	485	1	HXKA YEAST
16	64	11.6	320	1	RLX1 STAU
17	64	11.6	1278	1	MDR2 BAR
18	64	11.6	1609	1	LMG1 HUMAN
19	63.5	11.5	166	1	RAG2 ORYSA
20	63.5	11.5	513	1	Y4W1 RHISN
21	63.5	11.5	828	1	MRKC KLEPN
22	63.5	11.5	1019	1	LFC CARRO
23	63.5	11.5	1019	1	LFC CARRO
24	63	11.4	159	1	LFC TACR
25	63	11.4	1279	1	MDR3 LYCES
26	62.5	11.3	235	1	MDR3 HUMAN
27	62.5	11.3	250	1	THM2 THADA
28	62.5	11.3	488	1	ATL1 LYCES
29	62	11.3	158	1	CRYP CRYPE
30	62	11.3	159	1	PRO6 LYCES
31	62	11.2	159	1	HSC6 DROME
32	61.5	11.2	455	1	YB12 YEAST
33	61	11.1	501	1	ATL1 ARATH

ALIGNMENTS

34	61	11.1	574	1	HMD3 SOLTU	Q41438 solanum tub
35	61	11.1	1276	1	MDR1 CRIGR	P21448 cricetus
36	61	11.1	1786	1	LMB1 HUMAN	P07942 homo sapien
37	61	11.1	1790	1	LMB1 DROME	P10466 drosophila
38	60.5	11.0	265	1	PPGK MYCTU	Q59568 mycobacteri
39	60.5	11.0	348	1	LLY LEGN	Q53407 legionella
40	60.5	11.0	403	1	PGK CHLNU	P94666 chlamydia t
41	60.5	11.0	692	1	ANR6 HUMAN	Q97294 homo sapien
42	60.5	11.0	985	1	DROL HSV1	P28859 ictaluriid h
43	60	10.9	259	1	EXP6 ARATH	Q38655 arabidopsis
44	60	10.9	335	1	G3P2 METAC	P58838 methanosarc
45	60	10.9	423	1	CBP2 WHEAT	P08819 triticum ae

AC	P80915: 004396;	STANDARD;	PRT;	102 AA.
DT	15-JUL-1998 (Rel. 36, Created)			
DT	15-JUL-1998 (Rel. 36, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Antimicrobial peptide 1 precursor (AMP1).			
OS	Macadamia integrifolia (Macadamia nut).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OX	Spermatophyta; Magnoliophyta; eudicotyledons; Proteaceae; Macadamia.			
NCBI_Taxid=60698;				
[1]				
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 27-96.			
RC	TISSUE=Seed;			
RA	MEDLINE=97261828; PubMed=9108242;			
RX	Marcus J.P., Green J.L., Goulter K.C., Harrison S.J., Mannere J.M.,			
RT	"Purification, Characterisation and cDNA cloning of an antimicrobial			
RL	peptide from Macadamia integrifolia."			
RL	Eur. J. Biochem. 244:743-749(1997).			
CC	-1- FUNCTION: ANTIMICROBIAL PEPTIDE WHICH INHIBITS THE GROWTH OF			
CC	A VARIETY OF FUNGI, OOMYCETES, GRAM-POSITIVE BACTERIAL			
CC	PHYTOPATHOGENES AND S.CEREVISIAE IN VITRO. NO ACTIVITY AGAINST			
CC	E.COLI.			
CC	-1- MISCELLANEOUS: ITS ANTIMICROBIAL ACTIVITY IS DIMINISHED BY CALCIUM			
CC	AND POTASSIUM CHLORIDE SALTS.			
CC	CC This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib.ch).			
DR	EMBL: Y10903; CA71842.1; -			
KW	Plant defense; Antibiotic; Fungicide; signal.			
FT	SIGNAL			
FT	CHAIN 27 102			
SQ	SEQUENCE 102 AA; 10943 MW; 9FB79954454B9311 CRC64;			
Query Match	100.0%; Score 551; DB 1; Length 102;			
Best Local Similarity	100.0%; Pred. No. 1.0e-51;			
Matches 102; Conservative	0; Mismatches 0; Indels 0; Gaps 0;			
QY	1 MASTKLFPSVITVMMILMAAEMVNGSAFTVWSGCGCNRRARERYSKCGCSA1HOKGVDF 60			
DB	1 MASTKLFPSVITVMMILMAAEMVNGSAFTVWSGCGCNRRARERYSKCGCSA1HOKGVDF 60			
QY	61 STTGQTALYNQAGCGVAHTRFGSSARACNPFMGKSIPIQC 102			
DB	61 STTGQTALYNQAGCGVAHTRFGSSARACNPFMGKSIPIQC 102			
RESULT 2				
HXKB_YEAST				

FT TURN 165 166
 FT STRAND 169 169
 FT TURN 175 176
 FT STRAND 187 187
 FT HELIX 188 199
 FT STRAND 202 208
 FT HELIX 210 221
 FT TURN 223 224
 FT STRAND 225 231
 FT STRAND 235 241
 FT HELIX 243 245
 FT TURN 251 252
 FT STRAND 263 266
 FT TURN 270 275
 FT HELIX 283 291
 FT HELIX 299 304
 FT HELIX 306 322
 FT TURN 323 324
 FT TURN 338 339
 FT TURN 343 343
 FT HELIX 344 351
 FT HELIX 358 368
 FT TURN 369 369
 FT HELIX 374 395
 FT TURN 395 396
 FT HELIX 397 406
 FT TURN 407 407
 FT STRAND 411 416
 FT TURN 418 421
 FT TURN 424 425
 FT HELIX 426 438
 FT HELIX 445 447
 FT STRAND 450 454
 FT TURN 458 460
 FT HELIX 461 475
 SQ SEQUENCE 485 AA; 53811 MW; E4A7EC796125324E CRC64;

Query Match 13.7%; Score 75.5; DB 1; Length 485;
 Best Local Similarity 30.2%; Pred. No. 1;
 Matches 29; Conservative 9; Mismatches 27; Indels 31; Gaps 6;

QY 8 FSVITVWMM---LIMASEMNGSAFTVWSGPCNNRAERYSKGCSAIIHOKGSDPSYSGT 64
 DB 368 FGIINTVVOEKRLIRRLSELGA-----RAARLSVGIATICQKRGK---TG 411
 QY 65 QTAALYNQAGCSGVATRF--GSSARACNP---FGW 95
 DB 412 HIAA-----DGSVYRYPGPFKEKANKALKDLYGW 440

RESULT 3
 ID_FGK_CHLMU STANDARD; PRT; 403 AA.
 AC 09PLN4;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Phosphoglycerate kinase (EC 2.7.2.3).
 GN PGK OR TC0065.
 OS Chlamydia muridarum.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 NC NCB1_TaxID=83560;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MOPn / N199;
 RA MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 White O., Hickey E.K., Peterson J., Uetebach T., Berry K., Bass S.,
 Linher K., Weisman J., Knout H., Craven B., Bowman C., Dodson R.,
 RA Guin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis MOPn and Chlamydia
 pneumoniae AR39.";

RL Nucleic Acids Res. 28:1397-1406(2000).
 CC -1- CATALYTIC ACTIVITY: ATP + 3-phospho-D-glycerate = ADP + 3-
 CC phospho-D-glyceroyl phosphate.
 CC -1- PATHWAY: Second phase of glycolysis; second step.
 CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE PHOSPHOGLYCERATE KINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE002274; AAF73528.1; -
 CC HSRP: P36204; IYDE.
 DR TIGR: TC0065; -
 DR InterPro: IPR001576; PGK.
 DR Pfam: PF00162; PGK; 1.
 DR PRINTS: PRO0477; PHGLYCKINASE.
 DR PROSITE: PS00111; PGlycerate KINASE; 1.
 KW Transferase; kinase; Glycolysis; Complete proteome.
 SQ SEQUENCE 403 AA; 43391 MW; 7A80C7A55089F64 CRC64;

Query Match 13.0%; Score 71.5; DB 1; Length 403;
 Best Local Similarity 29.3%; Pred. No. 2.3;
 Matches 24; Conservative 13; Mismatches 28; Indels 17; Gaps 4;

QY 17 IMASEMNGSAFTVWSG-----PGCNRAERYSKC--GCSAIIHOKGSDPSYSGT 66
 DB 307 IALFSEVIQDSATVPMNPGVGEVPPDQSKAIAQCLASHSAVTVWG-----GDA 360
 QY 67 AALYNQAGC-SGVATRFSSA 87
 DB 361 AAVVALAGCTSQISHVSTGGG 382

RESULT 4
 ID_MDR3_CRIGR STANDARD; PRT; 1281 AA.
 AC P23174;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Multidrug resistance protein 3 (P-glycoprotein 3).
 GN PGY3 OR PGP3.
 OS Cricetus griseus (Chinese hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Cricetulus.
 NC NCB1_TaxID=10029;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=92135896; PubMed=1685679;
 RA Endicott J.A., Sarangi F., Ling V.;
 RT "Complete cDNA sequences encoding the Chinese hamster P-glycoprotein
 RT gene family";
 RL DNA Seq. 2:89-101(1991).
 CC -1- FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED
 CC DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- MISCELLANEOUS: PGP ISOFORMS DIFFER IN THEIR DRUG TRANSPORT
 CC CAPABILITIES: PGP1 AND PGP2 CAN MEDIATE MDR, WHILE PGP3 APPARENTLY
 CC CANNOT.
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

FT DISULFID 979 991 BY SIMILARITY.
FT DISULFID 981 998 BY SIMILARITY.
FT DISULFID 1000 1009 BY SIMILARITY.
FT DISULFID 1012 1023 BY SIMILARITY.
FT CARBOHYD 101 101 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 121 121 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 507 507 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 643 643 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 725 725 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 921 921 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1062 1062 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1168 1168 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1211 1211 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1244 1244 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1321 1321 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1437 1437 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1557 AA; 172723 MW; CAF0B51F8D558DZF CRC64;

Query Match 12.4%; Score 68.5; DB 1; Length 1557;
Best Local Similarity 28.8%; Pred. No. 16;
Matches 19; Conservative 5; Mismatches 33; Indels 9; Gaps 2;

QY 29 FTWSPGCGNNRERYSKCGSAIHCKGYDFSVTGTALYNQAG--CSGVNHTFPGSS 86
Db 920 YNTSGLG-----QECNCDPLSGSGNTCVNTGCGQCKPVGTCRCDCADYHFGFS 972
QY 87 ARACNP 92
Db 973 ANCGP 978

RESULT 6
MDR1 HUMAN
ID MDR1_HUMAN STANDARD; PRT; 1280 AA.
AC P08183; O12755; O14812;
DT 01-ANG-1988 (Rel. 08, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Multidrug resistance protein 1 (P-glycoprotein 1) (CD243 antigen).
GN ABCB1 OR PGY1 OR MDR1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN 1
RP SEQUENCE FROM N.A.
RX MEDLINE=87028230; PubMed=2876781;
RA Chen C.-J., Chin J.E., Ueda K., Clark D.P., Pastan I., Gottesman M.M.,
RA Roninson I.B.;
RT "Internal duplication and homology with bacterial transport proteins
RT in the mdr1 (P-glycoprotein) gene from multidrug-resistant human
RT cells";
RL Cell 47:381-389(1986).
RN 12
RP SEQUENCE FROM N.A.
RX MEDLINE=90094448; PubMed=1967175;
RA Chen C.-J., Clark D.P., Ueda K., Pastan I., Gottesman M.M.,
RA Roninson I.B.;
RT "Genomic organization of the human multidrug resistance (MDR1) gene
RT and origin of P-glycoproteins";
RL J. Biol. Chem. 265:506-514(1990).
RN 13
RP SEQUENCE FROM N.A.
RX MEDLINE=97190336; PubMed=9038218;
RA Chen G., Duran G.E., Steger K.A., Lacayo N.J., Jaffrezou J.P.,
RA Dumontet C., Sakic B.I.;
RT "Multidrug-resistant human sarcoma cells with a mutant P-glycoprotein,
RT altered phenotype, and resistance to cyclosporins";
RL J. Biol. Chem. 272:5974-5982(1997).
RN 14
RP SEQUENCE OF 1-234 FROM N.A.
RA Smith A., Beck C., Gibson A.;

RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 178-215 AND 800-856 FROM N.A.
RX MEDLINE=90290529; PubMed=1972623;
RA Gekeler V., Weger S., Probst H.;
RT "mdr1/P-glycoprotein gene segments analyzed from various human
RT leukemic cell lines exhibiting different multidrug resistance
RT profiles";
RL Biochem. Biophys. Res. Commun. 169:796-802(1990).
RN 16
RP SEQUENCE OF 1-23 FROM N.A.
RA Kikita N., Tsubota J., Kakehi Y., Komano T., Gottesman M.M.,
RA Pastan I., Ueda K.;
RT Submitted (Jul-1991) to the EMBL/GenBank/DBJ databases.
RN 17
RP VARIANTS SER-893 AND THR-893.
RX MEDLINE=21686803; PubMed=11829140;
RA Saito S., Iida A., Sekine A., Mura Y., Ogawa C., Kawachi S.,
RA Higuchi S., Nakamura Y.;
RT "Three hundred twenty-six genetic variations in genes encoding nine
RT members of ATP-binding cassette, subfamily B (ABCB/MDR/TAP), in the
RT Japanese population";
RL J. Hum. Genet. 47:38-50(2002).
CC - FUNCTION: ENERGY-DEPENDENT EFFLUX PUMP RESPONSIBLE FOR DECREASED
CC DRUG ACCUMULATION IN MULTIDRUG-RESISTANT CELLS.
CC - SUBCELLULAR LOCATION: Integral membrane protein.
CC - TISSUE SPECIFICITY: EXPRESSED IN LIVER, KIDNEY, SMALL INTESTINE
CC AND BRAIN.
CC - SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC - DATABASE: NAME=Atlas Gene. Cytogenet. Oncol. Haematol.;
CC WWW="http://www.infobiogen.fr/services/chronocancer/genes/pgy1id105.html".
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; M14758; AAA59575.1; -;
DR EMBL; M29447; AAA59576.1; -;
DR EMBL; M29424; AAA59576.1; JOINED.
DR EMBL; M29425; AAA59576.1; JOINED.
DR EMBL; M29426; AAA59576.1; JOINED.
DR EMBL; M29427; AAA59576.1; JOINED.
DR EMBL; M29428; AAA59576.1; JOINED.
DR EMBL; M29429; AAA59576.1; JOINED.
DR EMBL; M29430; AAA59576.1; JOINED.
DR EMBL; M29431; AAA59576.1; JOINED.
DR EMBL; M29432; AAA59576.1; JOINED.
DR EMBL; M29433; AAA59576.1; JOINED.
DR EMBL; M29434; AAA59576.1; JOINED.
DR EMBL; M29435; AAA59576.1; JOINED.
DR EMBL; M29436; AAA59576.1; JOINED.
DR EMBL; M29437; AAA59576.1; JOINED.
DR EMBL; M29438; AAA59576.1; JOINED.
DR EMBL; M29439; AAA59576.1; JOINED.
DR EMBL; M29440; AAA59576.1; JOINED.
DR EMBL; M29441; AAA59576.1; JOINED.
DR EMBL; M29442; AAA59576.1; JOINED.
DR EMBL; M29443; AAA59576.1; JOINED.
DR EMBL; M29444; AAA59576.1; JOINED.
DR EMBL; M29445; AAA59576.1; JOINED.
DR EMBL; M29446; AAA59576.1; JOINED.
DR EMBL; AF016535; AAB69423.1; -;
DR EMBL; AC002457; AAC82531.1; -;
DR EMBL; M37724; AAA88047.1; -;
DR EMBL; M37725; AAA88048.1; -;
DR EMBL; X58723; CAA41558.1; -;
DR PIR; A25059; DVH01.
DR PIR; A34914; A34914.
DR Genew; HGNC:40; ABCB1.

```

DR 1M1050:
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003439; ABC_transporter.
DR InterPro: IPR001140; ABC_transporter_TM.
DR Pfam: PF00005; ABC_tran; 2.
DR Pfam: PF00664; ABC_membrane; 2.
DR ProDom: PD000006; ABC_transporter; 2.
DR SMART: SM00382; AAA; 2.
DR PROSITE: PS00211; ABC_TRANSPORTER; 2.
DR ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
KW Multigene family; Polymorphism.
FT DOMAIN 1 51 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 52 72 POTENTIAL.
FT TRANSSEM 120 140 POTENTIAL.
FT TRANSSEM 189 209 POTENTIAL.
FT TRANSSEM 216 236 POTENTIAL.
FT TRANSSEM 297 317 POTENTIAL.
FT TRANSSEM 326 346 POTENTIAL.
FT TRANSSEM 347 710 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 711 731 POTENTIAL.
FT TRANSSEM 757 777 POTENTIAL.
FT TRANSSEM 833 853 POTENTIAL.
FT TRANSSEM 854 874 POTENTIAL.
FT TRANSSEM 937 957 POTENTIAL.
FT TRANSSEM 974 994 POTENTIAL.
FT DOMAIN 995 1280 CYTOPLASMIC (POTENTIAL).
FT NP_BIND 427 434 ATP (BY SIMILARITY).
FT NP_BIND 1070 1077 ATP (BY SIMILARITY).
FT REPEAT 1 637
FT REPEAT 638 1280
FT CARBOHYD 91 91 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 99 99 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 893 893 /FTID=VAR_013361.
FT VARIANT 893 893 A -> T.
FT VARIANT 893 893 S -> R (IN REF. 6).
FT CONFLICT 23 23 G -> V (IN REF. 1 AND 3).
FT CONFLICT 185 185 MISSING (IN REF. 3).
FT CONFLICT 336 336 G -> A (IN REF. 3).
FT CONFLICT 412 412 G -> S (IN REF. 3).
FT CONFLICT 438 438
FT SEQUENCE 1280 AA; 141462 MW; ABIC279531F43675 CRC64;

Query Match 12.3%; Score 68; DB 1; Length 1280;
Best Local Similarity 24.5%; Pred. No. 15;
Matches 24; Conservative 13; Mismatches 33; Indels 28; Gaps 3;

QY 6 LFFSVITVMILIAASEMV-----NGSAFTVWSGPCNNRAERYSKGCSAITHQKGYD 59
Db 334 VFFSVLIGAFVSGQSPSIEAFANARGAAVEIFKIIDNKPSIDSYSKSHKPNKGNLE 393
QY 60 -----FSY-----TGQTAALYNQAGC 75
Db 394 FRNVHFSYPSRKEVKILKGLNLKVQSGQTVALVGNSGC 431

RESULT 7
VEF_GVHA
ID_VEF_GVHA STANDARD; PRT; 902 AA.
AC P54232;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Viral enhancing factor (VEF) (Enhancin) (104 kDa glycoprotein)
DE (Synergistic factor).
GN VEF.
OS Heliothis armigera granulosis virus (HaGV) (Heliothis armigera granulovirus).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae; Granulovirus.
OX NCBI_TaxID=45440;
RN [1]
RP SEQUENCE FROM N.A.

QY 26 GSAP--TVWSGPCNNRAE--RYSKGCSTAIHQK-GYDFSYTGOT--AALYNQAGCGV 78
Db 213 GGAYYGAFMTAPASTNLGEYLRVSTNMWVIHELGHAYDFVFTVNTRLIEIWNNSFCDRI 272
QY 79 AHTRFSSAR 88
Db 273 QYTNWNTKR 282

RESULT 8
ID_CSGB_SALTY STANDARD; PRT; 151 AA.
AC P55226;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Minor curlin subunit precursor (Fimbrin SEF17 minor subunit).
GN CSGB OR AGFB OR STM1143.
OS Salmonella typhimurium, and
OS Salmonella enteritidis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602, 592;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=SR-11;
RX MEDLINE=98117058; PubMed=9457880;
RA Romling U., Bian Z., Hammar M., Sierralta W.D., Normark S.;
RT "Curli fibers are highly conserved between Salmonella typhimurium and
RT Escherichia coli with respect to operon structure and regulation.";
RL J. Bacteriol. 180:722-731(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,

```

RA Courtney L., Portojlik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan B., Sun H., Flores L., Miller W., Stoneking T., Nan M.,
 RA Waterston R., Wilson R.K.,
 RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium
 LT7.";
 RL Nature 413:852-856(2001).
 RP [3]
 RN SEQUENCE FROM N.A.
 RC SPECIES=*S. enteritidis*; STRAIN=27655-3B;
 RX MEDLINE=96146512; PubMed=8550497;
 RA Collinson S.K., Clouthier S.C., Doran J.L., Baner P.A., Kay W.W.,
 RT "Salmonella enteritidis *agfBAC* operon encoding thin, aggregative
 RT fimbriae.";
 RL J. Bacteriol. 178:662-667(1996).
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
 CC CURLIN MONOMERS.
 CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC -----
 CC CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AJ002301; CAA05316.1; -
 DR EMBL: AE008749; AAL20073.1; -
 DR EMBL: U43280; AAC43598.1; -
 DR StryGene; SG10609; csgb.
 KW Fimbria; Signal; Complete proteome.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
 SQ SEQUENCE 151 AA; 16182 MW; C0FC5430E6DD361D CRC64;
 Query Match 12.1%; Score 66.5; DB 1; Length 151;
 Best Local Similarity 31.0%; Pred. No. 3.3;
 Matches 18; Conservative 11; Mismatches 16; Indels 13; Gaps 2;
 QY 30 TWMSGPCNNRAERYKSCGSAIHQKGYDFSYTGQTALYNQAGCGVAHTFGSSA 87
 DB 69 SVISQEGNNRAK-----VDQAGNYPAYIEQT-----GNANDASISQAYGNSA 113
 RESULT 9
 RA05_ORYSA STANDARD; PRT; 157 AA.
 ID RA05_ORYSA
 AC Q01861;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last annotation update)
 DE Seed allergenic protein RA5 precursor.
 GN RA5.
 OS *Oryza sativa* (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhacoidae; Oryzaceae; Oryza.
 NCBI_Taxid=4530;
 RX MEDLINE=93144699; PubMed=7678765;
 RA Adachi T., Izumi H., Yamada T., Tanaka K., Takeuchi S.,
 RA Nakamura R., Matsuda T.,
 RT "Gene structure and expression of rice seed allergenic proteins
 RT belonging to the alpha-amylase/trypsin inhibitor family.";
 RL Plant Mol. Biol. 21:239-248(1993).
 CC -1- PTM: FIVE DISULFIDE BONDS ARE PRESENT (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE CEREAL TRYPSIN/ALPHA-AMYLASE INHIBITOR

CC FAMILY.
 CC -----
 CC CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D11430; BAA01996.1; -
 DR PIR: S31078; S31078.
 DR HSSP: P01085; IHSS.
 DR InterPro; IPR003612; AAI.
 DR InterPro; IPR001768; Try/amy1 inhbr.
 DR Pfam; PF00234; tryp alpha amyl. 1.
 DR PRINTS; PR00808; AMLASEINSTR.
 DR SMART; SM00499; AAI. 1.
 DR PROSITE; PS00426; CEREAL_TRYP_AMYL_INH. 1.
 KW Allergen; Multigene family; Signal.
 FT SIGNAL 1 26 POTENTIAL.
 FT CHAIN 27 157 SEED ALLERGENIC PROTEIN RA5.
 SQ SEQUENCE 157 AA; 17118 MW; C0A5495FFB399E6 CRC64;
 Query Match 12.1%; Score 66.5; DB 1; Length 157;
 Best Local Similarity 22.1%; Pred. No. 3.4;
 Matches 33; Conservative 12; Mismatches 35; Indels 69; Gaps 7;
 QY 1 MASTKLFPSV--ITVMMLIMASEVNVN-----GSAFTWMSGPCNNRAE 42
 DB 1 MASNKVFESVLLLAVALAVATATMAEYHODQVYTRACOPGMYSLPRRALVK 60
 QY 43 RY-----SKGCSAI-HQKGYDFSYTGQTALYNQAGCGS 77
 DB 61 RQCRGSAAAAEQVRDCCQQLAAVDSWCRCEATSHMLGCG-----TYRELGAPD 109
 QY 78 VAH-----TRFGSAPA-----CN 91
 DB 110 VGHPMSEVFRGCRGDLERAAALPAPCN 138
 RESULT 10
 RA17_ORYSA STANDARD; PRT; 162 AA.
 ID RA17_ORYSA
 AC Q01863;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 28, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last annotation update)
 DE Seed allergenic protein RA17 precursor.
 GN RA17.
 OS *Oryza sativa* (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhacoidae; Oryzaceae; Oryza.
 NCBI_Taxid=4530;
 RX MEDLINE=93144699; PubMed=7678765;
 RA Adachi T., Izumi H., Yamada T., Tanaka K., Takeuchi S.,
 RA Nakamura R., Matsuda T.,
 RT "Gene structure and expression of rice seed allergenic proteins
 RT belonging to the alpha-amylase/trypsin inhibitor family.";
 RL Plant Mol. Biol. 21:239-248(1993).
 CC [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Seed;
 RX MEDLINE=92289999; PubMed=1376283;
 RA Izumi H., Adachi T., Fujii N., Matsuda T., Nakamura R., Tanaka K.,
 RA Urisu A., Kurosawa Y.,
 RT "Nucleotide sequence of a cDNA clone encoding a major allergenic
 RT protein in rice seeds. Homology of the deduced amino acid sequence
 RT with members of alpha-amylase/trypsin inhibitor family.";

OC Salmonella.
 OX NCBI_TaxID=601;
 RN [1]
 RC SEQUENCE FROM N.A.
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Felwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,
 RA Krogan A., Larsen T.S., Leather S., Moute S., O'Gea P., Parry C.,
 RA Quail M., Rutherford K., Simmonds S., Skelton J., Stevens K.,
 RA Whitehead S., Barrett B.G.;
 RT "Complete genome sequence of a multiple drug resistant Salmonella
 RT enterica serovar Typhimurium";
 RL Nature 413:848-852(2001)
 CC -1- FUNCTION: CURLIN IS THE STRUCTURAL SUBUNIT OF THE CURLI. CURLI ARE
 CC COILED SURFACE STRUCTURES THAT ASSEMBLE PREFERENTIALLY AT GROWTH
 CC TEMPERATURES BELOW 37 DEGREES CELSIUS. CURLI CAN BIND TO
 CC FIBRONECTIN. THE MINOR SUBUNIT IS THE NUCLEATION COMPONENT OF
 CC CURLIN MONOMERS.
 CC -1- SIMILARITY: BELONGS TO THE CSGA/CSGB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AL627269; CAD0267.1; -
 KW Fibria; Signal; Complete proteome.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 151 MINOR CURLIN SUBUNIT.
 SQ SEQUENCE 151 AA; 16254 MW; 161C54326E573495 CRC64;
 Query March 11.9%; Score 65.5; DB 1; Length 151;
 Best Local Similarity 31.0%; Pred. No. 4.2;
 Matches 18; Conservative 11; Mismatches 16; Indels 13; Gaps 2;
 QY 30 TWGSPGNNRRAERYSKGCSAIIHQGGYDFSYGTQTAALYNQAGCGVAHTFGSSA 87
 DB 69 SVISGEGNNRAK-----VDAQNYNFAVIEQT-----GNANDASISQSAVGNSSA 113
 RESULT 13
 OSLS ARATH STANDARD; PRT; 244 AA.
 AC PS0700; O970D2;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Osmotin-like protein OSM34 precursor.
 GN OSM34 OR AT4G11650 OR T5C23.80.
 OS Arabidopsis thaliana (Mouse-ear cress).
 CC Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC Eucosydes II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia; TISSUE=leaf;
 RA MEDLINE=97354294; PubMed=9210588;
 RA Capelli N., Diagon T., Greppin H., Simon P.;
 RT "Isolation and characterization of a cDNA clone encoding an
 RT osmotin-like protein from Arabidopsis thaliana.";
 RL Gene 191:51-56(1997).
 RP [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=20083488; PubMed=10617198;

RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
 RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
 RA Harris B., Anstorge W., Brandt P., Grivell L.A., Rieger M.,
 RA Weichselgartner M., de Simone V., Obermaier B., Maché R., Mueller M.,
 RA Kreis M., Delaeny M., Puigdomenech P., Watson M., Schichtel T.,
 RA Reinhardt B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
 RA Vos P., Hodelsel U., Zimmermann W., Wedler H., Ridley P.,
 RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
 RA Van der Schueren J., Grymoprez B., Chuang Y.-J., Vandenbusche F.,
 RA Braeken M., Weijens J., Voet M., Bastiens I., Aert R., Defoor E.,
 RA Weitzneger T., Bohe G., Ransperger U., Hilbert H., Braun M.,
 RA Holzer E., Brandt A., Peters S., van Stevenen M., Dikse W.,
 RA Woolfman P., Klein Lankhorst R., Rose M., Hauf U., Koeter P.,
 RA Bernieris S., Hempel S., Feldpausch M., Landwehr S., Van den Daele H.,
 RA De Keyser A., Buysnaert C., Gielen J., Villarroel R., De Clercq R.,
 RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
 RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,
 RA Petrecci A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
 RA Borrova D., Bioecker H., Scharfe M., Grimm M., Loehner T.-H.,
 RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fartmann B., Grandenath K., Danner D., Herzl A.,
 RA Neumann S., Argitlou A., Vitale D., Liguori R., Piravandi E.,
 RA Massenet O., Quigley F., Clabaud G., Mendenhall A., Felber R.,
 RA Schnabl S., Hiller R., Schmidt W., Lechery A., Aubourg S.,
 RA Cherdorf F., Cooke R., Berger C., Montfort A., Casacuberta E.,
 RA Gibbons T., Weber N., Vandenbol M., Barques M., Terol J., Torres A.,
 RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
 RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,
 RA Friehman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
 RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
 RA Parnell L., Dedha N., Gnoj L., Schnitz K., Huang E., Spiegel L.,
 RA Sekhon M., Murray J., Sheel P., Cordes M., Abu-Threideh J.,
 RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
 RA Lattelle P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
 RA Minx P., Bulten D., Fulton B., Miller N., Greco T., Kemp K.,
 RA Kramer U., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
 RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
 RA Du H., Ali J., Bergthoff A., Jones K., Drone K., Cotton M., Joshi C.,
 RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Jordan C.,
 RA Ma P., Zhong J., Preston R., Vil D., Shaker M., Matero A., Shah R.,
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,
 RA Grant S., Shohdy N., Hasegawa A., Hamed A., Lochi M., Johnson A.,
 RA Chen E., Maria M., Martienssen R., McCombie W.R.;
 RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
 RT thaliana.";
 RL Nature 402:769-777(1999).
 CC -1- SIMILARITY: BELONGS TO THE THAUMATIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X89008; CAA61411.1; -
 DR EMBL: AL049500; CAB3936.1; -
 DR EMBL: AL161532; CAB78208.1; -
 DR HSSP: P25871; 1A0N.
 DR InterPro: IPR001938; Thaumatin.
 DR Pfam: PF00314; Thaumatin.1.
 DR PRINTS: PRO0347; THAUMATIN.
 DR ProDom: PD001321; Thaumatin.1.
 DR SMART: SM00205; THN.1.
 DR PROSITE: PS00316; THAUMATIN.1.
 KW Signal.
 FT SIGNAL 1 22
 FT CHAIN 23 244
 FT DISULFID 31 222
 FT DISULFID 72 82
 FT DISULFID 87 93
 FT DISULFID 138 212
 POTENTIAL.
 OSMOTIN-LIKE PROTEIN OSM34.
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.

```

FT DISULFID 143 195 BY SIMILARITY.
FT DISULFID 151 161 BY SIMILARITY.
FT DISULFID 165 174 BY SIMILARITY.
FT DISULFID 175 182 BY SIMILARITY.
FT CONFLICT 186 186 E -> V (IN REF. 1).
SQ SEQUENCE 244 AA; 26633 MW; 9FBE9A45E9E195E0 CRC64;

Query Match 11.9%; Score 65.5; DB 1; Length 244;
Best Local Similarity 23.1%; Pred. No. 6.5;
Matches 24; Conservative 19; Mismatches 44; Indels 17; Gaps 3;

QY 1 MASTKLFFSVITVWMLIAMASEMWN--SGPCNNRAERYSKGCSAIHQKGGY 58
D 5 LVSTFISALLIISTATATTEILNQCSTYVAAASPGGRRLD-----AGQSRL 55
D 59 DFSYTGTAALYNOAGCGVAHTRFGSSARACNPFQWKSIFQIC 102
D 56 DVAAGTKMARIWGRNCRN-----FDSSGRGRCQTGDCSGGLQC 93

RESULT 14
ID RA14_ORYSA STANDARD; PRT; 165 AA.
AC Q01882;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Seed allergenic protein RA14 precursor.
GN Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Seed;
RX MEDLINE=93144699; PubMed=7678765;
RA Adachi T., Izumi H., Yamada T., Tanaka K., Takeuchi S.,
RA Nakamura R., Matsuda T.;
RT "Gene structure and expression of rice seed allergenic proteins
PL Plant Mol. Biol. 21:239-248(1993).
CC -1- PWM: FIVE DISULFIDE BONDS ARE PRESENT (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CEREAL TRYPSIN/ALPHA-AMYLASE INHIBITOR
CC FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sb-sib.ch).
CC -----
DR ENBL; D11432; BAA01998.1; -.
DR PIR; S31080; S31080.
DR HSP; P01085; IHSS.
DR InterPro; IPR003612; AAI.
DR InterPro; IPR001768; Try/alpha1 inhbr.
DR Pfam; PF00234; tryp_alpha_aml_1.
DR PRINTS; PR00808; AMLASEINHBR.
DR SMART; SW00499; AAI; 1.
DR PROSITE; PS00426; CEREAL_TRYP_AMYL_INH; 1.
KW Allergen; Multigene family; Signal.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 165 SEED ALLERGENIC PROTEIN RA14.
SQ SEQUENCE 165 AA; 17887 MW; 13F99783726CEFC6A CRC64;

Query Match 11.8%; Score 65; DB 1; Length 165;
Best Local Similarity 20.4%; Pred. No. 5.1;
Matches 31; Conservative 17; Mismatches 32; Indels 72; Gaps 7;

```

```

QY 1 MASTKLFFS--VITVWMLIAMASEMWN-----GSAFTWSPGCGNNRAE 42
D 1 MASNKVVSALLIIIVSLAATRMADHKKDQVYSLGERCPQGMGYSLPRCAVVK 60
QY 43 RY-----SKGCSAI-HOKGGYDFSYTGQTAALYNOAG 74
D 61 RQCVGTRSPGAVDEQLAQCCRELAADVDSWCRCSALNHMVG-----IYRELG 109
QY 75 CSGVAH-----TRFGSSARA-----CN 91
D 110 ATDVGHFPAEVPFGCRGDLERAAASLPAFCN 141

RESULT 15
HXKA YEAST
ID HXKA YEAST STANDARD; PRT; 485 AA.
AC P04806;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hexokinase A (EC 2.7.1.1) (Hexokinase PI).
GN HXK1 OR HKA OR YPR053C.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86120382; PubMed=3003701;
RA Stachel C., Stachel J., Swan J., Botstein D., Konigsberg W.;
RT "Identification, cloning and sequence determination of the genes
RT specifying hexokinase A and B from yeast.";
RL Nucleic Acids Res. 14:945-963(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86083199; PubMed=3908224;
RA Kopecki E., Entian K.-D., Mecke D.;
RT "Complete nucleotide sequence of the hexokinase PI gene (HXK1) of
RT Saccharomyces cerevisiae.";
RL Gene 39:95-102(1985).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=95400292; PubMed=7670463;
RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
RA Yamazaki M., Tashiro H., Eki T.;
RT "Analysis of the nucleotide sequence of chromosome VI from
RT Saccharomyces cerevisiae.";
RL Nat. Genet. 10:261-268(1995).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=96287652; PubMed=9686379;
RA Eki T., Naitou M., Hagiwara H., Ozawa M., Sasanuma S.-I.,
RA Sasanuma M., Tsuchiya Y., Shibata T., Hanaoka F., Murakami Y.;
RT "Analysis of a 36.2 kb DNA sequence including the right telomere of
RT chromosome VI from Saccharomyces cerevisiae.";
RL Yeast 12:149-167(1996).
RN [5]
RP ATP-BINDING AND SEQUENCE OF 104-112.
RX MEDLINE=88227998; PubMed=3131329;
RA Tamura J.K., Ladime J.R., Cross R.L.;
RT "The adenine nucleotide binding site on yeast hexokinase PII.
RT Affinity labeling of Lys-111 by pyridoxal
RT 5'-diphospho-5'-adenosine.";
RL J. Biol. Chem. 263:7907-7912(1988).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS).
RX MEDLINE=81049624; PubMed=7001031;
RA Bennett W.S. Jr., Steitz T.A.;
RT "Structure of a complex between yeast hexokinase A and glucose. I.
RT Structure determination and refinement at 3.5-A resolution.";

```

```

RL J. Mol. Biol. 140:183-210(1980).
CC -1- CATALYTIC ACTIVITY: ATP + D-hexose = ADP + D-hexose 6-phosphate.
CC -1- ENZYME REGULATION: SUBJECT TO ALLOSTERIC CONTROL. SUBSTRATE
CC -1- INHIBITION BY ATP.
CC -1- PATHWAY: FIRST STEP OF SEVERAL METABOLIC PATHWAYS.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBUNIT: HOMODIMER.
CC -1- MISCELLANEOUS: IN YEAST THERE ARE THREE GLUCOSE-PHOSPHORYLATING
CC -1- ISOMERISOMES, DESIGNATED HEXOKINASE I, II AND GLUCOKINASE.
CC -1- SIMILARITY: BELONGS TO THE HEXOKINASE FAMILY.
CC -1- DATABASE: NAME=Worthington Biochem manual;
CC WWW="http://www.worthington-biochem.com/manual/H/HK.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M14410; AAA34698.1; -.
DR EMBL; X03482; CAA27202.1; -.
DR EMBL; D50617; BAA09292.1; -.
DR PIR; A24531; KIBYHA.
DR PIR; A28178; A28178.
DR PDB; 1HKJ; 15-OCT-91.
DR SWISS-2DPAGE; P04806; YEAST.
DR SGD; S0001949; HXK1.
DR InterPro; IPR001312; Hexokinase.
DR Pfam; PF00349; hexokinase; 1.
DR PRINTS; PR00475; hexokinase2; 1.
DR PRODOM; PD001109; hexokinase; 1.
DR PROSITE; PS00378; hexokinases; 1.
DR Transferrase; Kinase; Glycolysis; Allosteric enzyme; ATP-binding;
KW 3D-structure.
KM
FT BINDING 111 111
FT DOMAIN 152 178
FT CONFLICT 61 61
FT CONFLICT 103 103
FT CONFLICT 194 194
FT CONFLICT 244 244
FT CONFLICT 356 357
FT CONFLICT 364 364
FT CONFLICT 388 388
FT CONFLICT 444 444
FT CONFLICT 479 480
FT TURN 19 20
FT HELIX 21 29
FT HELIX 32 33
FT HELIX 38 44
FT TURN 45 46
FT HELIX 47 54
FT TURN 55 56
FT STRAND 66 67
FT STRAND 82 82
FT STRAND 85 85
FT STRAND 88 88
FT TURN 89 90
FT STRAND 91 92
FT STRAND 94 94
FT STRAND 97 97
FT STRAND 111 111
FT TURN 115 119
FT STRAND 121 122
FT TURN 124 126
FT TURN 128 130
FT HELIX 131 140
FT HELIX 142 143
FT STRAND 161 161
FT STRAND 171 171
FT TURN 176 177
FT STRAND 180 181

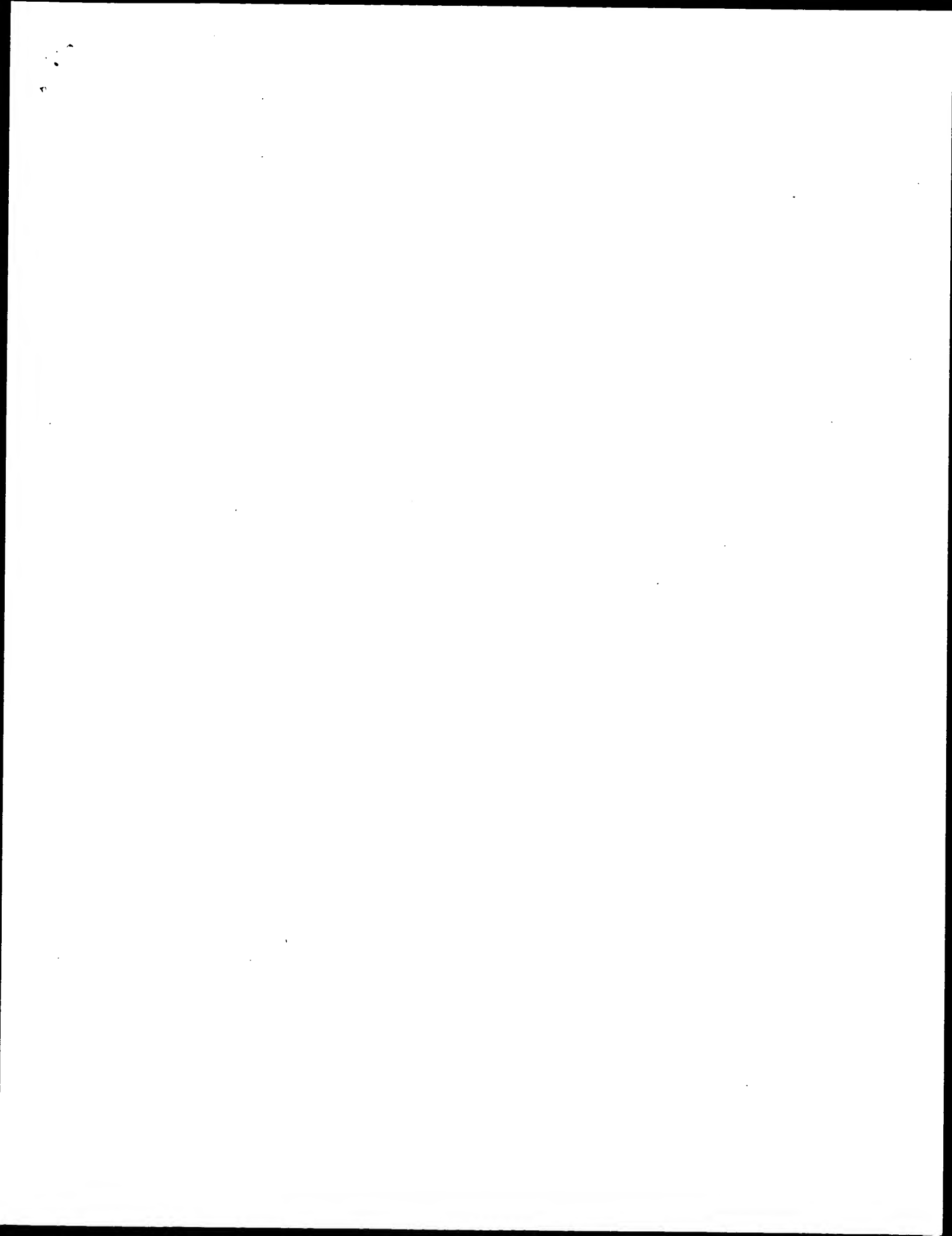
```

```

FT TURN 189 195
FT HELIX 196 200
FT TURN 211 216
FT HELIX 218 222
FT TURN 224 225
FT STRAND 226 232
FT STRAND 236 239
FT STRAND 242 242
FT TURN 244 245
FT STRAND 254 264
FT STRAND 256 267
FT HELIX 270 272
FT TURN 286 287
FT HELIX 288 292
FT TURN 293 294
FT HELIX 300 303
FT TURN 304 305
FT TURN 308 310
FT HELIX 311 323
FT TURN 324 325
FT TURN 339 340
FT TURN 345 351
FT TURN 359 369
FT HELIX 375 377
FT TURN 378 396
FT TURN 398 401
FT HELIX 402 406
FT TURN 407 408
FT STRAND 412 417
FT TURN 419 420
FT HELIX 428 438
FT STRAND 450 450
FT STRAND 454 454
FT TURN 458 460
FT HELIX 461 474
FT TURN 475 477
SQ SEQUENCE 485 AA; 53738 MM; AF5C9DA8F17BC3D0 CRC64;
Query Match 11.7%; Score 64.5; DB 1; Length 485;
Best Local Similarity 27.4%; Pred. No. 15;
Matches 17; Conservative 9; Mismatches 19; Indels 17; Gaps 2;
QY 40 RAERYKCGSAIHQKQYDFSTTGTALYN-----QKCGVATRTFGSSARACNPF 93
Db 391 RAARLAVCGIALICQKRGYTGHTADGSVYKYPGFKKAARGLR-----DIY 439
QY 94 GW 95
Db 440 GW 441

```

Search completed: January 12, 2003, 09:36:39
 Job time : 27 secs



GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2003, 09:30:05 ; Search time 80 Seconds

(without alignments)
262.710 Million cell updates/sec

Title: US-09-882-434A-1

Sequence: 1 MASTKLFPSVITVMMLIAMA.....FGSSARACNPFMKSLFIQC 102

Scoring table: BIOSUM62

Gap 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mhc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP virus:*
13: SP vertebrate:*
14: SP unclassified:*
15: SP rvirus:*
16: SP bacteriophage:*
17: SP archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	77.5	14.1	486	3	Q05838 saccharomyc
2	75.5	13.7	1558	5	Q9BLR5 Q9BLR5 leishmania
3	75	13.6	254	5	Q45025 hydra magni
4	71	12.9	730	2	Q07667 enterococcu
5	70	12.7	307	12	Q9WQ79 felina infe
6	70	12.7	307	12	Q9WQ80 canine coro
7	69.5	12.6	532	16	Q8YD38 bruceella me
8	69.5	12.6	1792	13	Q57484 gallus galli
9	69	12.5	668	6	Q8WNU6 mus musculu
10	69	12.5	1280	6	Q46605 canis fami
11	68	12.3	712	11	Q8VHD9 canis fami
12	67.5	12.3	184	5	Q09943 caenorhabd
13	67	12.2	307	12	Q9WQ75 felina infe
14	67	12.2	653	12	Q9IBR0 spodoptera
15	67	12.2	1486	4	Q14637 homo sapien
16	67	12.2	2121	12	Q9YMB7 transmissib

17	67	12.2	6685	12	Q9IM06	Q9IM06 transmissib
18	66.5	12.1	81	10	Q24225	Q24225 oryza sativ
19	66.5	12.1	251	16	Q9ZBHS	Q9ZBHS streptomyc
20	66.5	12.1	342	4	Q96BE4	Q96BE4 homo sapien
21	66.5	12.1	461	16	Q8ZDL1	Q8ZDL1 yeastsia pe
22	66.5	12.1	996	3	Q8X001	Q8X001 neurospora
23	66	12.0	78	10	Q8W4V6	Q8W4V6 capsaicin an
24	66	12.0	111	5	Q95ZS9	Q95ZS9 caenorhabd
25	66	12.0	210	10	Q949L8	Q949L8 beta vulgar
26	66	12.0	387	12	Q8Q9R0	Q8Q9R0 chimpanzee
27	66	12.0	534	3	Q74996	Q74996 felis silve
28	66	12.0	1163	6	Q9TSU2	Q9TSU2 schizosacch
29	65.5	11.9	197	3	Q9P7F1	Q9P7F1 caenorhabd
30	65.5	11.9	233	5	Q45533	Q45533 ralsomonia
31	65.5	11.9	382	16	Q8XT05	Q8XT05 mus musculu
32	65.5	11.9	580	5	Q9IUB8	Q9IUB8 dictyosteli
33	65.5	11.9	2346	11	Q9IUC1	Q9IUC1 xenopus lae
34	65	11.8	199	5	Q94476	Q94476 ratius norv
35	65	11.8	256	10	Q8S290	Q8S290 ratius norv
36	65	11.8	766	10	Q8W3Y6	Q8W3Y6 ratius norv
37	65	11.8	1086	4	Q75808	Q75808 ratius norv
38	64.5	11.7	147	4	Q9H5U7	Q9H5U7 ratius norv
39	64.5	11.7	1007	13	Q90ZM3	Q90ZM3 ratius norv
40	64	11.6	54	5	Q9BK84	Q9BK84 mesobuthus
41	64	11.6	225	10	Q8S4P7	Q8S4P7 tritium ae
42	64	11.6	767	13	Q9DGR2	Q9DGR2 xenopus lae
43	64	11.6	1272	11	Q9JG64	Q9JG64 ratius norv
44	64	11.6	1272	11	Q924L6	Q924L6 ratius norv
45	64	11.6	1551	16	Q8YMU3	Q8YMU3 anabena sp

ALIGNMENTS

RESULT 1
Q05838
ID Q05838
AC Q05838; PRELIMINARY; PRT; 486 AA.
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hexokinase PI (HKX2).
GN HKX2 OR YGL253W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86083199; PubMed=3908224;
RA Kopeckzi E., Entian K.D., Mecke D.;
RT "Complete nucleotide sequence of the hexokinase PI gene (HKX1) of
RT Saccharomyces cerevisiae.";
RL Gene 39:95-101(1985).
DR EMBL, M14411; AAA34699.1; ..
DR HSSP; P19367; IOHA.
DR SGP; S0003222; HKX2.
DR InterPro; IPR001312; Hexokinase.
DR Pfam; PF00349; hexokinase2.1.
DR Pfam; PF03727; hexokinase2.1.
DR PRINTS; PR00475; HEXOKINASE.
DR ProDom; PD001109; Hexokinase; 1.
DR PROSITE; PS00378; HEXOKINASES; 1.
SQ SEQUENCE 486 AA; 53952 MW; 218FAC9B8C36C8 CRC64;

Query Match 14.1%; Score 77.5; DB 3; Length 486;
Best Local Similarity 31.2%; Pred. No. 2;
Matches 30; Conservative 8; Mismatches 27; Indels 31; Gaps 6;
QY 8 FSVITVM---LIMASEMVNGSAFTWSPGCGNRRARYSKCCGSAIHOXGDFSYTG 64
DB 369 FQINTVGERLIRLRLSLIGA-----RAARLSVCGIAICQKRGK---TG 412

RX MEDLINE=9319897; PubMed=10392726;
 RA Stephensen C.B., Casebolt D.B., Gangopadhyay N.N.;

RT "Phylogenetic analysis of a highly conserved region of the polymerase
RT gene from 11 coronaviruses and development of a consensus polymerase
chain reaction assay.";

RL Virus Res. 60:181-189(1999).

DR EMBL: AF124987; AAD32991.1; -.

KW RNA-directed RNA polymerase.

FT NON TER 1 307 307

SQ SEQUENCE 307 AA; 34473 MW; 2A3AD4C643AC6E1 CRC64;

Query Match 12.7%; Score 70; DB 12; Length 307;

Best Local Similarity 25.0%; Pred. No. 8.3;

Matches 26; Conservative 13; Mismatches 35; Indels 30; Gaps 4;

DB 16 L1AMASEMWSAFTWSSGCGNNRABRYSKG-----CSA1HKGYSYTGOTAA 68

DB 124 MIRNASAMTIGSKHV-----GCCTHSDFRRLSNELAOVLTEVHCTGCGFYKGGTTSG 178

QY 69 LYNQAGSGVAHTRF-----GSSARACNPFQWKS1 98

DB 179 ----DGTAVANSAFNIFQAVSANVKKLGVDSNACNNVTVKS1 218

RESULT 6

Q9WQ80 PRELIMINARY; PRT; 307 AA.

AC Q9WQ80; (TREMBlrel. 12, Created)

DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)

DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)

DE RNA-directed RNA polymerase (Fragment).

OS POL.

OC canine coronavirus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;

OC Coronaviridae; Coronavir.

OX NCBI_TaxID=11153;

RN [1]

RC SEQUENCE FROM N.A.

RA STRAIN=1-71;

RX MEDLINE=99319897; PubMed=10392726;

RA Stephenson C.B., Casbolt D.B., Gangopadhyay N.N.;

RT "Phylogenetic analysis of a highly conserved region of the polymerase
RT gene from 11 coronaviruses and development of a consensus polymerase
chain reaction assay.";

RL Virus Res. 60:181-189(1999).

DR EMBL: AF124986; AAD32990.1; -.

KW RNA-directed RNA polymerase.

FT NON TER 1 307 307

SQ SEQUENCE 307 AA; 34500 MW; 2A3AD4C643AC9711 CRC64;

Query Match 12.7%; Score 70; DB 12; Length 307;

Best Local Similarity 25.0%; Pred. No. 8.3; Indels 30; Gaps 4;

DB 16 L1AMASEMWSAFTWSSGCGNNRABRYSKG-----CSA1HKGYSYTGOTAA 68

DB 124 MIRNASAMTIGSKHV-----GCCTHSDFRRLSNELAOVLTEVHCTGCGFYKGGTTSG 178

QY 69 LYNQAGSGVAHTRF-----GSSARACNPFQWKS1 98

DB 179 ----DGTAVANSAFNIFQAVSANVKKLGVDSNACNNVTVKS1 218

RESULT 7

Q8YD38 PRELIMINARY; PRT; 532 AA.

AC Q8YD38; (TREMBlrel. 20, Created)

DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)

DT 01-MAR-2002 (TREMBlrel. 21, Last annotation update)

DE Alcohol dehydrogenase (acceptor) (EC 1.1.99.-).

GN BME110339.

OS Brucella melitensis.

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Brucellaceae; Brucella.

OX NCBI_TaxID=29459;

RN [1]

RC SEQUENCE FROM N.A.

RA STRAIN=16W / ATCC 23456 / B10TYPE 1;

RX MEDLINE=20020109; PubMed=11756688;

RA Delvecchio V.G., Kapatal V., Redkar R.J., Patra G., Mujer C., Los T.,

RA Ivanova N., Anderson I., Bhattacharya A., Lykdis A., Reznik G.,

RA Jablonki L., Larsen N., D'Souza M., Bernal A., Mazur M., Coleman E.,

RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-D.,

RA Haselkorn R., Kyrides N., Overbeek R.,

RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).

DR EMBL: AE009671; AAL53581.1; -.

DR InterPro: IPR00172; GMC_OXRED.

DR Pfam: PF0732; GMC_OXRED_1.

DR PROSITE: PS00624; GMC_OXRED_2; 1.

KW Oxidoreductase; Complete proteome.

SQ SEQUENCE 532 AA; 58225 MW; F15B57645D149ADD CRC64;

Query Match 12.6%; Score 69.5; DB 16; Length 532;

Best Local Similarity 27.2%; Pred. No. 17; Indels 35; Gaps 4;

Matches 22; Conservative 8; Mismatches 16; Indels 35; Gaps 4;

DB 29 FTWSSGCGNNRABRYSK-----CGCSA1H-----QKGGYDSYTGOTAA LYNQAGSG 77

DB 64 FTTEAPGLNGRSLNYPKRGVLGGSSINGMTYMGQACDVG-----LMRQAGCGG 114

QY 78 VAHTRFGSSARACNPFQWKS1 98

DB 115 -----WGMDDV 120

RESULT 8

ID 057484 PRELIMINARY; PRT; 1792 AA.

AC 057484;

DT 01-JUN-1998 (TREMBlrel. 06, Created)

DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Laminin beta 2-like chain.

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

OX NCBI_TaxID=9031;

RN [1]

RC SEQUENCE FROM N.A.

RX MEDLINE=93015947; PubMed=1400373;

RA O'Rear J.J.;

RT "A novel laminin B1 chain variant in avian eye.";

RL J. Biol. Chem. 267:20555-20557(1992).

RN [2]

RC SEQUENCE FROM N.A.

RX MEDLINE=98209634; PubMed=9550264;

RA Liu J., Swaidson S., Xie W., Brewton R.G., Mayne R.;

RT "Primary structure and expression of a chicken laminin beta chain:

RT evidence for four beta chains in birds.";

RL Matrix Biol. 16:471-481(1998).

DR EMBL: AF038555; AAB92586.1; -.

DR HSSP: P02468; IKLO.

DR InterPro: IPR002106; AACRNA_ligaseII.

DR InterPro: IPR000561; EGF-like.

DR InterPro: IPR002049; Laminin_EGF.

DR InterPro: IPR001886; LamNT.

DR Pfam: PF00053; Laminin_EGF; 13.

DR Pfam: PF00055; Laminin_Nterm; 1.

DR PRINTS: PR00011; EGF_LAMININ.

DR SMART: SM00180; EGF_Lam; 13.

DR SMART: SM00136; LamNT; 1.

```

DR PROSITE; PS00339; AA TRNA LIGASE II 2; UNKNOWN_1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_10.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01248; LAMININ TYPE_EGF; 12.
KW EGF-like domain; Glycoprotein; Laminin EGF-like domain; Repeat.
SQ SEQUENCE 1792 AA; 195723 MW; 4A4CBE0206F6A600 CRC64;

Query Match 12.6%; Score 69.5; DB 13; Length 1792;
Best Local Similarity 25.0%; Pred. No. 61;
Matches 25; Conservative 9; Mismatches 41; Indels 25; Gaps 4;

QY 18 AMASEMVNGSAFTVWSPGCGN-NRAERY-----SKGCSAIHOKGGYD 59
DQ 300 ATAGVMVHRCVCKHHTQGLNCERCDFFHELPWRPAEGSSINACRDCNHSRCHFD 359
DQ 60 ---FVSVTGTAALYNOAGSGVAHTFGSSARACNPFPGWK 96
DQ 360 MAVFLATGNTSG----AVDCGQHHTMGRHCHLCKPFFYK 395

RESULT 9
Q8WNU6 PRELIMINARY; PRT; 668 AA.
AC Q8WNU6;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Multidrug resistance p-glycoprotein (fragment).
GN MDRI.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]_TaxID=9615;
RP SEQUENCE FROM N.A.
SQ SEQUENCE 668 AA; 73546 MW; F8CD80F7C934B31A CRC64;

Query Match 12.5%; Score 69; DB 6; Length 668;
Best Local Similarity 25.5%; Pred. No. 24;
Matches 25; Conservative 12; Mismatches 33; Indels 28; Gaps 3;

QY 6 LFFSVITVWMLIAMASEMV-----NGSAFTVWSPGCGNNAERYSKGCSAIHOKGGYD 59
DQ 85 VFFSVLIGAFSGQASPIEAFANARGAAYEIFKIIDNKPSIDSYSKSGHKPDNIKGNLE 144
QY 60 -----FSY-----TGQTAALYNOAGC 75
DQ 145 FKNVHFSYPSRKEVKILKGLNLKVGSGQTVALVGNSGC 182

RESULT 10
O46605 PRELIMINARY; PRT; 1280 AA.
AC O46605;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Multidrug resistance p-glycoprotein.
GN MDRI.

```

```

OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]_TaxID=9615;
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Puel O., Lepage J.F., Alvinerie M., Galtier P., Pineau T.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
CC -|- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL: AF045016; AAC02113.1;
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR001140; ABCtransportrTM.
DR InterPro; IPR003439; ABC transportr.
DR Pfam; PF00664; ABC membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR ProDom; PD000006; ABC transportr; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW ATP-binding; Transport.
SQ SEQUENCE 1280 AA; 141524 MW; 762DD5AFF4C73306 CRC64;

Query Match 12.5%; Score 69; DB 6; Length 1280;
Best Local Similarity 25.5%; Pred. No. 48;
Matches 25; Conservative 12; Mismatches 33; Indels 28; Gaps 3;

QY 6 LFFSVITVWMLIAMASEMV-----NGSAFTVWSPGCGNNAERYSKGCSAIHOKGGYD 59
DQ 335 VFFSVLIGAFSGQASPIEAFANARGAAYEIFKIIDNKPSIDSYSKSGHKPDNIKGNLE 394
QY 60 -----FSY-----TGQTAALYNOAGC 75
DQ 395 FKNVHFSYPSRKEVKILKGLNLKVGSGQTVALVGNSGC 432

RESULT 11
Q8VHD9 PRELIMINARY; PRT; 712 AA.
AC Q8VHD9;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Diversin.
GN ANKRD6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_TaxID=10090;
RP SEQUENCE FROM N.A.
RA Schwarz-Romond T., Asbrand C., Bakkers J., Kuhl M., Schaeffer H.-J.,
RA Huelken J., Behrens J., Hammerschmidt M., Birchmeier W.,
RT "The Ankyrin Repeat Protein Diversin Acts in the Wnt Signaling Pathway
RT and Controls Embryonic Axis Formation."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY026320; AAK15806.1;
DR MGI; MGI:2154278; Ankrd6.
DR InterPro; IPR002110; ANK.
DR Pfam; PF00023; ank; 8.
DR PRINTS; PR01415; ANKYRIN.
DR SMART; SM00248; ANK; 8.
DR PROSITE; PS50088; ANK_REPEAT; 6.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
SQ SEQUENCE 712 AA; 77934 MW; BC0344DA54E731A CRC64;

Query Match 12.3%; Score 68; DB 11; Length 712;
Best Local Similarity 23.5%; Pred. No. 34;
Matches 24; Conservative 21; Mismatches 37; Indels 20; Gaps 3;

QY 1 MASTKLPFSVITVWMLIAMASEMVNGSAFTVWSPGCGNNAERYSKGCSAIHOKGGYD 45
DQ 48 LAANKGHLSVVOILLKAGCDLVDQDGDQDTALHRTAVVGNTEILTALIREGC--ALDRQD 105

```


QY 43 RYKCGCSAI-----HOKGGYD-----PSYTGQTAAALYNQAGC---SGVAHTFGS 85
Db 65 KFSKRGCSMTCPFFHETGPDIDANTPANTQTSFTAILYAQPACYNLDRVAATREGA 121

RESULT 15

014637 PRELIMINARY; PRT; 1486 AA.
AC 014637;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Laminin alpha 3b chain (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=98055454; PubMed=9395076;
RA Doliana R., Bellina I., Buccioti F., Mongiat M., Perris R.,
RA Colombatti A.;
RT "The human alpha3b is a 'full-sized' laminin chain variant with a more
RT widespread tissue expression than the truncated alpha3a."
RL FEBS Lett. 417:65-70(1997).
DR EMBL; AF005258; AAC51867.1; .
DR HSPF; P02468; 1TLE.
DR InterPro; IPR000561; EGF-like
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001886; LamNT.
DR Pfam; PF00053; laminin EGF; 10.
DR Pfam; PF00055; laminin Nterm; 1.
DR PRINTS; PR00011; EGF_LAMININ.
DR PRODOM; PD002082; LamNT; 1.
DR SMART; SM00180; EGF_Lam; 10.
DR SMART; SM00001; EGF_Like; 1.
DR SMART; SM00136; LamNT; 1.
DR PROSITE; PS00022; EGF 1; UNKNOWN 10.
DR PROSITE; PS01248; LAMININ TYPE EGF; 9.
KW Glycoprotein; Laminin EGF-like domain; Repeat.
FT NON_TER 1
FT NON_TER 1486
SQ SEQUENCE 1486 AA; 162496 MW; F2E5EB7EB522E98C CRC64;

Query Match 12.2%; Score 67; DB 4; Length 1486;
Best Local Similarity 28.1%; Pred. No. 95;
Matches 27; Conservative 8; Mismatches 23; Indels 38; Gaps 7;

QY 34 GPGCNN-RAERYS-----KCGCSAIHQKGYDF---SYTGO-----TAAL 69
Db 480 GPRCDTCRSGFSFPTCQACWCSAL---GSYQMPCCSVTGOCECRPGVTGQRCRCLSGA 536
QY 70 YNQAGCGVAHTFGSSARACNPFG---WKSIFIOC 102
Db 537 YDFPHCOG-----SSSACDPAGTINWNLGYCQC 564

Search completed: January 12, 2003, 09:38:22
Job time : 84 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 12, 2003, 07:06:54 ; Search time 2665 Seconds

(without alignments)
5383.746 Million cell updates/sec

Title: US-09-882-434A-2

Perfect score: 493

Sequence: 1 attaagcttcgtgagctccat.....ccatgcatgatttacc 493

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:*

1: gb_da:*

2: gb_hcg:*

3: gb_in:*

4: gb_cm:*

5: gb_ov:*

6: gb_pac:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_rts:*

12: gb_ey:*

13: gb_un:*

14: gb_vi:*

15: gb_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_ov:*

22: em_pac:*

23: em_ph:*

24: em_pl:*

25: em_ro:*

26: em_rts:*

27: em_un:*

28: em_vi:*

29: em_ba:*

30: em_hcg_hum:*

31: em_hcg_inv:*

32: em_hcg_other:*

33: em_hcg_mus:*

34: em_hcg_pln:*

35: em_hcg_rtd:*

36: em_hcg_vtc:*

37: em_hcg_vtc:*

38: em_ey:*

39: em_hgo_hum:*

40: em_hgo_mus:*

41: em_hgo_other:*

Prod. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	492.6	99.9	493	8	MIRNAPI
2	41.2	8.4	100919	9	AL359378
3	39.4	8.0	168575	9	AC091989
4	39.2	8.0	110913	2	CNS07YQ2
5	39.2	8.0	137936	2	CNS08CAX
6	39.2	8.0	144778	2	CNS07EPR
7	39	7.9	152137	2	AP001932
8	39	7.9	153485	2	AC090757
9	39	7.9	175032	9	AC019240
10	38.6	7.8	99862	2	AC106935
11	38.6	7.8	149767	2	AC113238
12	38.4	7.8	105470	2	AC116306
13	38.4	7.8	159919	9	AL359194
14	38.2	7.7	80828	5	AL672211
15	38.2	7.7	126118	9	AC104598
16	38.2	7.7	222454	2	AC098159
17	38.2	7.7	248360	2	AC099351
18	38	7.7	120832	2	CNS08C9W
19	38	7.7	128417	2	CNS08C44
20	38	7.7	167473	2	AC118111
21	37.8	7.7	140929	2	AC117805
22	37.2	7.5	224	8	AY024154
23	37.2	7.5	443	8	HSP26C11
24	37.2	7.5	29855	8	AP003826
25	37.2	7.5	192833	2	AL672160
26	37.2	7.5	204104	2	AL732613
27	36.8	7.5	114964	10	AL691507
28	36.8	7.5	177032	9	CNS06C7W
29	36.8	7.5	213564	2	AC095604
30	36.8	7.5	236980	10	AL591478
31	36.8	7.5	304322	2	AC118787
32	36.6	7.4	125020	9	AF429315
33	36.4	7.4	385	11	HS024XB8
34	36.4	7.4	76245	2	AC125151
35	36.4	7.4	110000	2	AC125102
36	36.4	7.4	150655	9	AC025887
37	36.4	7.4	157979	9	AC009835
38	36.4	7.4	166408	2	AC102255
39	36.4	7.4	167941	9	AC092638
40	36.4	7.4	170082	2	AC129774
41	36.4	7.4	195296	2	AC124448
42	36.4	7.4	266229	2	AC073676
43	36.2	7.3	68147	2	AC123627
44	36.2	7.3	69201	9	AC004109
45	35.2	7.3	85907	8	ATF28M11

ALIGNMENTS

RESULT 1
LOCUS MIRNAPI
DEFINITION M.integrifolia mRNA for antimicrobial protein 1.
ACCESSION Y10903
VERSION Y10903.1 GI:2181942
KEYWORDS
SOURCE AMPI Gene.
ORGANISM Macadamia integrifolia.
Macadamia integrifolia.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; Proteaceae;
Macadamia.
REFERENCE
1 (bases 1 to 493)
Marcus,J.P., Goulter,K.C., Green,J.L., Harrison,S.J. and
Manners,J.M.

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
JOURNAL

Purification, characterisation and cDNA cloning of an antimicrobial peptide from *Macadamia integrifolia*
Eur. J. Biochem. 244 (3), 743-749 (1997)
97261828
2 (bases 1 to 493)
Marcos, J.P.
Direct Submission
Submitted (30-JAN-1997) J.P. Marcos, Cooperative Research Centre for Trop., Plant pathology., 5th Level John Hines Building., The University of Queensland., Brisbane, QLD 4072, AUSTRALIA
Location/Qualifiers
1. 493
/organism="Macadamia integrifolia"
/db_xref="taxon:60698"
/clone="pGM7*2"
/dev_stages="seed"
70. .378
/gene="AMPI"
70. .378
/gene="AMPI"
/function="antimicrobial protein"
/codon_start=1
/protein_id="CAA71842.1"
/db_xref="GI:2181943"
/db_xref="SWISS-PROT:P80915"
/translation="MASPKLFVITVIMLIAMASEMVNGSAFTVWSPGCGNRAERY SKCGSAIHQGGYDFSTGTGTAALYNQAGCGVATRFSSARACNPFQWKSIFIQ "

sig_peptide 70. .147
/gene="AMPI"
148. .375
/gene="AMPI"
/product="unnamed"
/evidence="experimental"

BASE COUNT 132 a 106 c 109 g 145 t 1 others
ORIGIN

Query Match 99.9%; Score 492.6; DB 8; Length 493;
Best Local Similarity 100.0%; Pred. No. 3.2e-121;
Matches 493; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTAAGTCTTTGAGTCTACATATCTCTTCTCCCTCCACCATTTAGCATTACAGCTA 60
Db 1 ATTAAGTCTTTGAGTCTACATATCTCTTCTCCCTCCACCATTTAGCATTACAGCTA 60
QY 61 ACCTGAGCCATGCTCCACCAAGTTGTTCTTCTCAGTCATTACTGTGATGCTCATTA 120
Db 61 ACCTGAGCCATGCTCCACCAAGTTGTTCTTCTCAGTCATTACTGTGATGCTCATTA 120
QY 121 GCAATGCAAGTGCAGATGGTGAATGGAGTGCAATTTACAGATATGGAGTGGTCCAGTTGT 180
Db 121 GCAATGCAAGTGCAGATGGTGAATGGAGTGCAATTTACAGATATGGAGTGGTCCAGTTGT 180
QY 181 AACACCGTCTGAGCGATATAGCAAGTGGATGTCTCAGTCATATACATCAGAGGGAGGC 240
Db 181 AACACCGTCTGAGCGATATAGCAAGTGGATGTCTCAGTCATATACATCAGAGGGAGGC 240
QY 241 TATGACTTCAGCTACACTGAGCAAACTGGCTCTCTACAAACAGGCTGGATCAGTGGT 300
Db 241 TATGACTTCAGCTACACTGAGCAAACTGGCTCTCTACAAACAGGCTGGATCAGTGGT 300
QY 301 GTTGACACACACAGGTTGGTCCAGTCCAGGCGATGCAACCCCTTTGGTTGGAAGAGT 360
Db 301 GTTGACACACACAGGTTGGTCCAGTCCAGGCGATGCAACCCCTTTGGTTGGAAGAGT 360
QY 361 ATCTTCATCCAACTGATGATATATCTCTGATCCATCTCTCTATGTTTTCAGAGTGT 420
Db 361 ATCTTCATCCAACTGATGATATATCTCTGATCCATCTCTCTATGTTTTCAGAGTGT 420
QY 421 ATAATTAGAGAGATGCATGATATATATAAATAAGTAAAGCTACGGTATCCCATGTG 480
Db 421 ATAATTAGAGAGATGCATGATATATAAATAAGTAAAGCTACGGTATCCCATGTG 480

QY 481 ATGATTTTACCC 493
Db 481 ATGATTTTACCC 493

RESULT 2

AL359378

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

DB 86357 AATAATCATGACTAAGTGAACCGTAATTAATTAAGTACTGTAATTAAGATTAATCT 86416
 QY 437 ATGATAT 444
 DB 86417 ATGGGTCT 86424

RESULT 5
 CNS08CAX 137936 bp DNA linear HTG 27-JUL-2002
 LOCUS Oryza sativa chromosome 12 clone OSJNBa0041K23, *** SEQUENCING IN
 DEFINITION PROGRESS ***, in ordered pieces.
 ACCESSION AL844874
 VERSION AL844874.1 GI:22003300
 KEYWORDS HTG; HTGS PHASE2; HTGS_ACTIVEFLN.
 SOURCE Oryza sativa.
 ORGANISM Oryza sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Eriacridae; Oryzaceae; Oryza.
 1 (bases 1 to 137936)
 Choinsene, N., Orjeda, G., Catolico, L., Demange, N., Wincker, P.,
 Segurens, B., Pelletier, E., Scarpelli, C., Salanoubat, M.,
 Weissenbach, J. and Quetier, F.
 Oryza sativa chromosome 12 sequencing
 Unpublished
 2 (bases 1 to 137936)
 Genoscope.
 Direct Submission
 Submitted (26-JUL-2002) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 Web : www.genoscope.cns.fr)
 IMPORTANT: This sequence is unfinished and does not necessarily
 represent the correct sequence.
 Work on the sequence is in progress and the release of this data is
 based on the understanding that the sequence may change as work
 continue. The sequence may be contaminated with foreign sequence
 from E.coli, yeast, vector, phage, etc.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES
 source
 1..137936
 /organism="Oryza sativa"
 /cultivar="Nipponbare"
 /sub_species="japonica"
 /db_xref="taxon:4530"
 /chromosome="12"
 /clone="OSJNBa0041K23"
 /clone_1ib="OSJNBa"
 /clone_1ib="OSJNBa"
 BASE COUNT 38960 a 29563 c 29308 g 40105 t
 ORIGIN

Query Match 8.0%; Score 39.2; DB 2; Length 137936;
 Best Local Similarity 50.5%; Pred. No. 5.6;
 Matches 95; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 257 CTGACAACTGCTGCTCTCTACCAACGAGCTGAGTGCAGTGTGTGACACACAGCT 316
 DB 119273 CTAACTACTACTACTCTACATATCTGTCAACCAATTCGCTCCAAATTCCTAAT 119214
 QY 317 TTGGGTCCAGTCCAGGCGATGCAACCCCTTTGGTTGAAGATATCTTCATCAATGCT 376
 DB 119213 TCGCTTTCTTCTAGAACTGCAAGCTGCTGCAGAAAAGAGATGATTAATTGGA 119154
 QY 377 AGATTCAATCTTGATTCATCTTCTATAGTTTTCAGAGTATTAATTAAGAGATGC 436
 DB 119153 AAAATCACTAGCTAAGTGAACCGTAATTAATTAAGTACTGTAATTAAGATATCT 119094
 QY 437 ATGATAT 444

DB 119093 ATGGGTCT 119086

RESULT 6
 CNS07EPR/c 144778 bp DNA linear HTG 06-FEB-2001
 LOCUS Oryza sativa chromosome 12 clone OSJNBa0041K23, *** SEQUENCING IN
 DEFINITION PROGRESS ***, in unordered pieces.
 ACCESSION AL513404
 VERSION AL513404.1 GI:12711302
 KEYWORDS HTG; HTGS PHASE1.
 SOURCE Oryza sativa.
 ORGANISM Oryza sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Eriacridae; Oryzaceae; Oryza.
 1 (bases 1 to 144778)
 Salse, J., Choinsene, N., Orjeda, G., Regad, F., Lorieux, M., Cooke, R.,
 Delser, M., Robert, C., Broctier, P., Wincker, P., Cruaud, C.,
 Artiguenave, F., Saurin, W., Salanoubat, M., Quetier, F. and
 Weissenbach, J.
 Oryza sativa chromosome 12 sequencing
 Unpublished
 2 (bases 1 to 144778)
 Genoscope.
 Direct Submission
 Submitted (06-FEB-2001) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 Web : www.genoscope.cns.fr)
 IMPORTANT: This sequence is unfinished and does not necessarily
 represent the correct sequence.
 Work on the sequence is in progress and the release of this data is
 based on the understanding that the sequence may change as work
 continue. The sequence may be contaminated with foreign sequence
 from E.coli, yeast, vector, phage, etc.
 Contigs composition :
 1062 bp contig from 1 to 1062
 933 bp contig from 2063 to 2995
 583 bp contig from 3996 to 4578
 2449 bp contig from 5579 to 8027
 66895 bp contig from 9028 to 75922
 67856 bp contig from 76923 to 144778.
 * NOTE: This is a 'working draft' sequence.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

FEATURES
 source
 1..144778
 /organism="Oryza sativa"
 /cultivar="Nipponbare"
 /sub_species="japonica"
 /db_xref="taxon:4530"
 /chromosome="12"
 /clone="OSJNBa0041K23"
 /clone_1ib="CUGi Nipponbare BAC"
 /clone_1ib="CUGi Nipponbare BAC"
 BASE COUNT 40362 a 30027 c 29723 g 39633 t 5033 others
 ORIGIN

Query Match 8.0%; Score 39.2; DB 2; Length 144778;
 Best Local Similarity 50.5%; Pred. No. 5.6;
 Matches 95; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 257 CTGACAACTGCTGCTCTCTACCAACGAGCTGAGTGCAGTGTGTGACACACAGCT 316
 DB 126114 CTAACTAGTAACTACTCAATATCTGTCAACCAATTCGCTCCAAATTCCTAAT 126055
 QY 317 TTGGGTCCAGTCCAGGCGATGCAACCCCTTTGGTTGAAGATATCTTCATCAATGCT 376
 DB 126054 TCGCTTTCTTCTAGAACTGCAAGCTGCTGCAGAAAAGAGATGATTAATTGGA 125995
 QY 377 AGATTCAATCTTGATTCATCTTCTATAGTTTTCAGAGTATTAATTAAGAGATGC 436
 DB 125994 AAAATCACTAGCTAAGTGAACCGTAATTAATTAAGTACTGTAATTAAGATATCT 125935

```

QY 437 ATGGATAT 444
|||||
Db 125934 ATGGGTCT 125927

RESULT 7
AP001932
LOCUS
DEFINITION
  Homo sapiens chromosome 18 clone RP11-701G2 map 18q22, WORKING
  DRAFT SEQUENCE, 6 unordered pieces.
ACCESSION
  AP001932
VERSION
  HTG; HTGS PHASE1; HTGS DRAFT.
KEYWORDS
  Homo sapiens DNA, clone:RP11-701G2.
SOURCE
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 152137)
    Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
    Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  Homo sapiens 152,137 genomic DNA of 18q22
  Published Only in DataBase (2000)
  2 (bases 1 to 152137)
    Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
    Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  Direct Submission
  Submitted (28-APR-2000) Masahira Hattori, The Institute of Physical
  and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
  Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
  Japan (E-mail:hattori@gscc.riken.go.jp,
  URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
  Fax:81-42-778-9924)
  On May 30, 2000 this sequence version replaced gi:7678794.
  ----- Genome Center
  Center: RIKEN Genomic Sciences Center (GSC)
  Center code: RIKEN
  Web site: http://hgp.gsc.riken.go.jp/
  Contact: hattori@gscc.riken.go.jp
  ----- Project Information
  Center project name: Humdraft18
  Center clone name: RP11-701G2
  ----- Summary Statistics
  Sequencing vector: PCR products; 100% of reads
  Chemistry: Dye-terminator ET-amersham; 100% of reads
  Assembly program: Phrap; version 0.990329
  Consensus quality: 146725 bases at least Q40
  Consensus quality: 149683 bases at least Q30
  Insert size: 151637; sum-of-contigs
  Quality coverage: 6.40x in Q20 bases; sum-of-contigs
  -----
  NOTE: This is a 'working draft' sequence. It currently consists of
  6 contigs. The true order of the pieces is not known and their
  order in this sequence record is arbitrary. Gaps between the
  contigs are represented as runs N, but the exact sizes of the gaps
  are unknown. This record will be updated with the finished sequence
  as soon as it is available and the accession number will be
  preserved
  1 34616 contig of 34616 bp in length
  34717 73705 contig of 38989 bp in length
  73806 103219 contig of 29414 bp in length
  103320 129829 contig of 26510 bp in length
  103930 144710 contig of 14781 bp in length
  144811 152137 contig of 7327 bp in length
  Sequence updated (26-May-2000).
  * NOTE: This is a 'working draft' sequence. It currently
  * consists of 6 contigs. The true order of the pieces
  * is not known and their order in this sequence record is
  * arbitrary. Gaps between the contigs are represented as
  * runs of N, but the exact sizes of the gaps are unknown.
  * This record will be updated with the finished sequence
  * as soon as it is available and the accession number will

```

```

* be preserved.
* 1 34616: contig of 34616 bp in length
* 34617 34716: gap of 100 bp
* 34717 73705: contig of 38989 bp in length
* 73706 73805: gap of 100 bp
* 73806 103219: contig of 29414 bp in length
* 103220 103319: gap of 100 bp
* 103320 129829: contig of 26510 bp in length
* 129830 129929: gap of 100 bp
* 129930 144710: contig of 14781 bp in length
* 144711 144810: gap of 100 bp
* 144811 152137: contig of 7327 bp in length.
FEATURES
  Location/Qualifiers
    1..152137
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="18"
      /map="18q22"
      /clone="RP11-701G2"
    misc_feature
      1..34616
        /note="assembly_fragment clone_end:77 vector_side:right"
    misc_feature
      34717..73705
        /note="assembly_fragment"
    misc_feature
      73806..103219
        /note="assembly_fragment clone_end:SP6 vector_side:left"
    misc_feature
      103320..129829
        /note="assembly_fragment"
    misc_feature
      129930..144710
        /note="assembly_fragment"
    misc_feature
      144811..152137
        /note="assembly_fragment"
  BASE COUNT 49625 a 26249 c 26522 g 49241 t 500 others
  ORIGIN
    Query Match 7.9%; Score 39; DB 2; Length 152137;
    Best Local Similarity 63.2%; Pred. No. 6.3;
    Matches 60; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
  QY 360 TATCTTCATCCAACTGCTAGATTCATACACTCTGGATCCATCTTCTATGTTTTCAAATG 419
  Db 59388 TATATACATACAAAGTATATATACATATATATATATATATATATATATATATATG 59447.
  QY 420 TATATTAGAGAGATGCTGGATATATATATATATATATATATATATATATATATATAT 454
  Db 59448 TATATATACATATATACATGATATATATATATATATATATATATATATATATATAT 59482
  RESULT 8
  AC090757
  LOCUS
  DEFINITION
    Homo sapiens chromosome 18 clone RP11-701G2 map 18, WORKING DRAFT
    SEQUENCE, 1 ordered piece.
  ACCESSION
    AC090757
  VERSION
    AC090757.2 GI:14192984
  KEYWORDS
    HTG; HTGS PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
  SOURCE
    Homo sapiens
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 153485)
    Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
    Barina,N., Bastien,V., Boguslavsky,L., Bouckhalter,B., Brown,A.,
    Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
    Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
    Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J.,
    Gargana,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
    Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
    Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
    Lechoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 153485)
    Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  Homo sapiens chromosome 18, clone RP11-701G2
  Unpublished
  2 (bases 1 to 153485)
    Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  Homo sapiens chromosome 18, clone RP11-701G2
  JOURNAL
  REFERENCE
  AUTHORS

```


Mon Jan 13 10:47:59 2003

Center project name: L1037
Center clone name: 510_D_19

FEATURES

source

Location/Qualifiers

1. .175032
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"

/clone="RP11-510D19"

/clone_lib="RPC1-11 Human Male BAC"

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

19381. .19425

/rpt_family="(TG)n"

complement(19438. .19517)

/rpt_family="L1MB1"

complement(19984. .20336)

/rpt_family="L1M4"

20331. .21039

/rpt_family="L1MB1"

21088. .21910

/rpt_family="L1MB1"

complement(21920. .22249)

/rpt_family="L1M4"

23063. .23442

/rpt_family="THE1C"

complement(23594. .23804)

/rpt_family="MIR"

23946. .23988

/rpt_family="AT_rich"

25330. .25351

/rpt_family="AT_rich"

25683. .25764

/rpt_family="GA-rich"

complement(26325. .26622)

/rpt_family="AluSg"

29475. .29670

/rpt_family="L1M4C"

29976. .30286

/rpt_family="AluSx"

complement(30681. .30961)

/rpt_family="L1PB3"

complement(31091. .31477)

/rpt_family="L1MDB"

31841. .31921

/rpt_family="MIR"

31921. .32143

/rpt_family="MIR"

32782. .32888

/rpt_family="L2"

complement(37388. .37743)

/rpt_family="MLT1A2"

37959. .38432

/rpt_family="MLT1B"

39629. .39664

/rpt_family="(CA)n"

complement(40423. .40659)

/rpt_family="MIR"

complement(41540. .41646)

/rpt_family="MIR"

42479. .42602

/rpt_family="(TA)n"

43338. .43362

/rpt_family="AT_rich"

complement(43384. .43690)

/rpt_family="AluJb"

43922. .43976

/rpt_family="AT_rich"

44458. .44754

Query Match 7.9%; Score 39; DB 9; Length 175032;

Best Local Similarity 63.2%; Pred.No. 6.2;

Matches 60; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

Qy 360 TATCTTCATCAATGCTAGATTTTCATAACTCTTGGATCCATCTTCTATGTTTTTCAAGTG 419

Db 123352 TATATACATACAGTATATATACATATATACATGATATATATATATATATATATATATG 123411

Qy 420 TATATATAGAGATGCGATGATATATAATAAATA 454

Db 123412 TATATACATATATACATGATATATAATAATA 123446

RESULT 10

AC106935

LOCUS AC106935 99862 bp DNA linear HTG 13-JUL-2002
 DEFINITION Rattus norvegicus clone CH230-134K22, *** SEQUENCING IN PROGRESS
 *** 47 unordered pieces.
 AC106935
 VERSION AC106935.2 GI:21736698
 KEYWORDS HTG; HTGS PHASE1.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 99862)
 REFERENCE
 AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C., Alshrook, S.L., Anaratunga, H.C., Are, U.R., Ayele, M., Banks, T., Barberia, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P., Buay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Caron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaery, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, P., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hult, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudan, S., Karlsson, E., Kelly, S., Khan, U., King, L., Koryan, J., Kovar, C., Krastevic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Liu, C., Liu, J., Liu, W., Louised, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mei, G., Metzner, M., Minter, G., Minter, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newson, N., Nguyen, A., Nguyen, S., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabot, P., Tamerisa, A., Tamerisa, K., Tang, H., Taney, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vaequez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorilla, S., Nelson, D., Weinstein, G., and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 99862)
 Worley, K.C.
 Direct Submission
 Submitted (14-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 99862)
 Worley, K.C.
 Direct Submission
 Submitted (13-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Jul 12, 2002 this sequence version replaced gi:18141352.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information

Center project name: COMU
 Center clone name: CH230-134K22
 ----- Summary Statistics
 Sequencing vector: Plasmid:
 Chemistry: Dye-terminator Big Dye 1004 of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 54814 bases at least Q40
 Consensus quality: 59343 bases at least Q30
 Consensus quality: 63485 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a "working draft" sequence. It currently
 * consists of 47 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 1307: contig of 1307 bp in length
 * 1308 1407: gap of unknown length
 * 1408 2662: contig of 1255 bp in length
 * 2663 2762: gap of unknown length
 * 2763 4238: contig of 1476 bp in length
 * 4239 4339: gap of unknown length
 * 4339 5561: contig of 1223 bp in length
 * 5562 6808: gap of unknown length
 * 6808 6908: contig of 1147 bp in length
 * 6908 8196: gap of unknown length
 * 8196 8297: contig of 1288 bp in length
 * 8297 9357: gap of unknown length
 * 9357 9457: gap of unknown length
 * 9457 10642: contig of 1185 bp in length
 * 10642 10741: gap of unknown length
 * 10741 11824: contig of 1083 bp in length
 * 11824 11925: gap of unknown length
 * 11925 13037: contig of 1113 bp in length
 * 13037 13137: gap of unknown length
 * 13137 14142: contig of 1005 bp in length
 * 14142 14242: gap of unknown length
 * 14242 15760: contig of 1518 bp in length
 * 15760 15860: gap of unknown length
 * 15860 17577: contig of 1717 bp in length
 * 17577 17677: gap of unknown length
 * 17677 19577: contig of 1900 bp in length
 * 19577 19677: gap of unknown length
 * 19677 20886: contig of 1209 bp in length
 * 20886 20986: gap of unknown length
 * 20986 22146: contig of 1160 bp in length
 * 22146 22246: gap of unknown length
 * 22246 24293: contig of 2047 bp in length
 * 24293 24393: gap of unknown length
 * 24393 25431: contig of 1038 bp in length
 * 25431 25531: gap of unknown length
 * 25531 27240: contig of 1709 bp in length
 * 27240 27340: gap of unknown length
 * 27340 28620: contig of 1280 bp in length
 * 28620 28720: gap of unknown length
 * 28720 30172: contig of 1452 bp in length
 * 30172 30272: gap of unknown length
 * 30272 31788: contig of 1516 bp in length
 * 31788 31888: gap of unknown length
 * 31888 34311: contig of 2423 bp in length
 * 34311 34411: gap of unknown length
 * 34411 36437: contig of 2026 bp in length
 * 36437 36537: gap of unknown length
 * 36537 38524: contig of 1987 bp in length
 * 38524 38624: gap of unknown length
 * 38624 39719: contig of 1095 bp in length
 * 39719 39820: gap of unknown length
 * 39820 40934: contig of 1115 bp in length

```

* 40935 41034: gap of unknown length
* 41035 42872: contig of 1838 bp in length
* 42873 42972: gap of unknown length
* 42973 44956: contig of 1984 bp in length
* 44957 45056: gap of unknown length
* 45057 48346: contig of 3290 bp in length
* 48347 48446: gap of unknown length
* 48447 51183: contig of 2737 bp in length
* 51184 51283: gap of unknown length
* 51284 53555: contig of 2272 bp in length
* 53556 53655: gap of unknown length
* 53656 56150: contig of 2495 bp in length
* 56151 56250: gap of unknown length
* 56251 59131: contig of 2881 bp in length
* 59132 59231: gap of unknown length
* 59232 61673: contig of 2442 bp in length
* 61674 61773: gap of unknown length
* 61774 64165: contig of 2392 bp in length
* 64166 64265: gap of unknown length
* 64266 66077: contig of 1812 bp in length
* 66078 66177: gap of unknown length
* 66178 68907: contig of 2730 bp in length
* 68908 69007: gap of unknown length
* 69009 72063: contig of 3056 bp in length
* 72064 72163: gap of unknown length
* 72164 74260: contig of 2097 bp in length
* 74261 74360: gap of unknown length
* 74361 78074: contig of 3714 bp in length
* 78075 78174: gap of unknown length
* 78175 80314: contig of 2140 bp in length
* 80315 80414: gap of unknown length
* 80415 83281: contig of 2867 bp in length
* 83282 83381: gap of unknown length
* 83382 87885: contig of 4404 bp in length
* 87886 87885: gap of unknown length
* 87886 91507: contig of 3622 bp in length
* 91508 91607: gap of unknown length
* 91608 94867: contig of 3260 bp in length
* 94868 94967: gap of unknown length
* 94968 99862: contig of 4895 bp in length.

FEATURES             Location/Qualifiers
     source            1..99862
                        /organism="Rattus norvegicus"
                        /db_xref="taxon:10116"
                        /clone="CH230-134K22"
BASE COUNT      27849 a 19699 c 19114 g 28488 t 4712 others
ORIGIN
Query Match      7.8%; Score 38.6; DB 2; Length 99862;
Best Local Similarity 32.0%; Pred. No. 8.3;
Matches 83; Conservative 1; Mismatches 175; Indels 0; Gaps 0;
QY 231 GAAGGAGGCTATGACTTCAGCTACACTGGACAACTGCTGCTCTTACACAGGCTGG 290
Db 53529 GAGGCGGAGCAAGAAAGGAGAGGTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 53588
QY 291 ATGCACTGGTGTGCACACACACACAGGTTTGGGTCAGTCCAGGCGCATGCAACCCCTTTTGG 350
Db 53589 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 53648
QY 351 TTGGAAGAGATATCTTCATCAATGCTAGATTTTCATCAACTCTTGGATCCATCTTCTATGTT 410
Db 53649 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 53708
QY 411 TTTCAGTGTATAATTTAGAGATGATGATGATATATAATAATAAGTAAAGTAAAGTAAAGTAA 470
Db 53709 TTTTATTTCAATCTCTATATACATATTTCAATATTTCAATATTTCAATATTTCAATATTTATGCT 53768
QY 471 TCACCATGTGATGATTTT 489
Db 53769 GCACATGGTCAAGCATGTT 53787

```

```

RESULT 11
AC113238
LOCUS
DEFINITION
AC113238
AC113238.1 GI:18958665
HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE
ORGANISM
Felis catus
REFERENCE
AUTHORS
Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Ho,S.-L., Idol,J.R., Karlins,E., Laric,P., Lee-Lin,S.-Q.,
Legaspi,R., Maduro,Q.L., Maduro,V.B., Mastello,C., Maskeri,B.,
Masirian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Prasad,A.,
Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W.,
Tsurgonov,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 149767)
Green,E.D.
Direct Submission
Submitted (27-FEB-2002) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.mouse@nhgri.nih.gov
----- Project Information
Center project name: cou
Center clone name: 252E08
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 147596 bases at least Q40
Consensus quality: 147747 bases at least Q30
Consensus quality: 147904 bases at least Q20
Insert size: 135000; agarose-fp
Quality coverage: 10.56x in Q20 bases; agarose-fp
Quality coverage: 9.55x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 9118: contig of 9118 bp in length
* 9119 9218: gap of unknown length
* 9219 17046: contig of 7828 bp in length
* 17047 17146: gap of unknown length
* 17147 34899: contig of 17753 bp in length
* 34900 34999: gap of unknown length
* 35000 57008: contig of 22008 bp in length
* 57008 57107: gap of unknown length
* 57108 149767: contig of 92660 bp in length.
Location/Qualifiers
1..149767
/organism="Felis catus"
/db_xref="taxon:9685"
/clone="RP86-252E8"
/clone_lib="RP86"
misc_feature 1..9118

```

```

AC113238      149767 bp      DNA      linear      HTG 27-FEB-2002
Felis catus clone RP86-252E8, WORKING DRAFT SEQUENCE, 5 unordered
pieces.
AC113238
AC113238.1 GI:18958665
HTG; HTGS_PHASE1; HTGS_DRAFT.
Felis catus
Felis catus
Felis catus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
1 (bases 1 to 149767)
Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Ho,S.-L., Idol,J.R., Karlins,E., Laric,P., Lee-Lin,S.-Q.,
Legaspi,R., Maduro,Q.L., Maduro,V.B., Mastello,C., Maskeri,B.,
Masirian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Prasad,A.,
Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W.,
Tsurgonov,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 149767)
Green,E.D.
Direct Submission
Submitted (27-FEB-2002) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.mouse@nhgri.nih.gov
----- Project Information
Center project name: cou
Center clone name: 252E08
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 147596 bases at least Q40
Consensus quality: 147747 bases at least Q30
Consensus quality: 147904 bases at least Q20
Insert size: 135000; agarose-fp
Quality coverage: 10.56x in Q20 bases; agarose-fp
Quality coverage: 9.55x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 9118: contig of 9118 bp in length
* 9119 9218: gap of unknown length
* 9219 17046: contig of 7828 bp in length
* 17047 17146: gap of unknown length
* 17147 34899: contig of 17753 bp in length
* 34900 34999: gap of unknown length
* 35000 57008: contig of 22008 bp in length
* 57008 57107: gap of unknown length
* 57108 149767: contig of 92660 bp in length.
Location/Qualifiers
1..149767
/organism="Felis catus"
/db_xref="taxon:9685"
/clone="RP86-252E8"
/clone_lib="RP86"
misc_feature 1..9118

```



```

/db_xref="GI:19807760"
/translation="MIFDSLSISLIPISMIMWGCCSNVSIELIMKQSQSHALVTFFQ
FATVAFINRWKQVSIWIGPIGLRERKIPKLTIFLMVSIIFILSVLNKALDC
DIPPHMIPRSSLLSTVIGFIFKYSQQILSLIMVTLGIIFATFSSMPDCK
EISLGHENLFRSIGMLMIAAMFLSSILGLIOETKYLKGRHYETIFYSVNDIL
HHQLQNSALMALPGFSGFPLWYLIVNVLVTOVTCIGVFIITGKTSTLCTLVI
SIRKLSIISVIFNNHFTSLTGTILVFLGTMYSTSGVKIEPLPTQVKEIE
KEKKLN"
complement(23858..24619)
/notes="ORF ID:dd_00658"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAL99330.1"
/db_xref="GI:19807761"
/translation="MKILITFILLINLIIVKSLVVKTVNYSPGCTGGVKNKVLHSH
CNLVFRELVMNVSYFANFKMDSCKNAKTVAEPLHVCDEDNQIKVSNYLTLYQ
DAINDLPMSCEVINGGCEENYIITSILKGTCSPIKYNDTEAMTYSTCDNSMLSY
ICSEPTFCIGNCAINPKYKNCQCVPGRLDGNADGIYKFPVKLAKQKPSPTP
SQTSTPDEDEPNHKSVPASIALIIFGLLSL"
complement(join(25977..30925,31064..31175))
/notes="ORF ID:dd_00657"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAL99331.1"
/db_xref="GI:19807762"
/translation="MILRLNPKPIITFLFILLVLLNENNVFSQQPYQTCSEWEPDV
GDMGFSVIASHPNISATSPKATSSVSFCTDDMGSSAYSLAGYIYAQFLYNESDPN
TOMTSLVOTVSLPILPNEQVLEYFLSQFETNINPVLNVYVDRPVASYNGNLLG
DLRVLPLESLPTVANIATTSYTNNGGGGGGGTQSGFRIGSVIIRKPAATWEE
VYVDVAKSGKYNGTVAAPFSIIQALSNVDFNGTLYLADCTYCNLSVLPISBIN
IISQSGDPTKCIISGEYQLNLFKISYNSQDDFLTSPRRGRFRQMESYFIDISNLI
TSISGSGSGSGSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS
GGAALFTGNVASFNCIFTSNADSNQNGSIARGNAALILENTLVQNCSCVASF
GPAISMLGFGSPPTISLVNCTMENILVDSNTPNMODSNIQIEPIEGALCSGI
MLRQSTQSTNDCIGGNGALCIISSNTSNTNFGNLSKSGPMGIYISQSSIV
TIQNVLEDNANGNIGAIIPAMELSKVNIFNTIIRNIGTNSLSSSDSLVLQSQIV
NTSNDYTIYLFSGKFLDSNINSGSFYATQTVISISLTFDDISGSHVPODQV
EITDSIFNLGFGFYSSKIHIVLSIENSITQFLMISPLNSVLISSVSSSDTDTI
INLLSQNLVNTAIGKFDNFSTNSILTQIEGSGSTFKNLFQNRSPFLGGCIF
TFSSVLDNCTAIGNFAKOGVSTIAGNLITIDGIIYONNSAGGGVIFYSTPG
PTIKNAIFSGNTASVGNNTADAALIDVEDPPSPSGQDTYSGRIYIKDEGOIV
SGFEVKIPLSGNTISTSTYNGIALFISNTLYXGPIINTVDVEFLSNPNLPVQTV
SNVLKSPGOIPISTKIECTEGRSYNGETCVQCPSIHAYCPGAKVQPEGWY
DVNQTPKPYLYECNPICISNTSCAPISFGPLCAQCDPGYDWMGECRCKDTS
PILFIAVYSGIMVLVQSDSSEGLMTVIIYFAQTLMLVSTGVKFNFLALNLFQ
PGSSSIFGICPGPDYQOYFTLLSAPMLIVLLISTVFFIVYVRYKLIQKLNKI
IKEQEDIEKSKSISVNSIIIVTSDPTINNLSLKSLERKLLRKKOQOQPLENDD
DSDFTSTTNDGDLRDSYTIADGGKHSILGKSIETISIGLSLQVTLIDNATAI
PSNQTFLSQGLGAFIKMLMNIYPIGKATFELFCENVGVSATVLIANGVSCYTDQY
HKAAMVYCLLIIVIGPLIIMILLFLNRKSLDDPHQRTYGVFIKYSFLVFDVI
LLRRLIILMTSTMDPTSAARSFLVSVLSVLLQKYQPKRISDRNLELTSLL
PICCVLNDTVQSIQEWIVSSSIIFFLHVVLYSNRYQIIDSLOHLYIISRGK
FICCVLNDTVQSIQEWIVSSSIIFFLHVVLYSNRYQIIDSLOHLYIISRGK
KGRKSNNTLPGSNSLYNQDHYNDDDPSRSGIFSKDKKQSDTKVKSINN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
DDEEDDNGESTNLLKP"
join(34837..35014,35188..35398,35512..36769)
/notes="ORF ID:dd_00653"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAL99332.1"
/db_xref="GI:19807763"
/translation="MSEITDNDIKKEMDNKNRDPDPTDNKELLKWKTKRKRKSNDSMHI
NNFYRNPNNFKLIANKDSNATKELRVLLDHDGLRIELPNLYCPTLTLRLNYL
WISQNLKILINDNDNKIIKIGIDIGTGTCTIFPLGAKLFNNWSFGIIDDIDKY
LEYAQNITINSKLTLPQNEKNSDILLKLANYPEKSNFTNSNDHDDHDDDD
DEYFADFCLNPPFKDLNENNNNNNNPKSNTCSVNMVTDGCEFEVKELIKES
FOLKCKRFTLTYGRKVLNPLNIIKQYLPKNOIQTITELVQNTSRWLSWFL
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NKSTNLETKENNNTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
IVGSSIKLDRDIEFLTFIDLTLLTFKLPIDPKINNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNSCDGEIINSNLFILKFLNLIKNEIKK"
37054..38214
/notes="ORF ID:dd_01574"

```

```

/codon_start=1
/product="PAP-1. 6/101"
/protein_id="AAL99333.1"
/db_xref="GI:19807764"
/translation="MDRRDYSKEKYKGYDNDKORRFSNDKIRHNFDEFTMPSNDRDR
DKDIDRYDRDRDRDRDRDRDRDRDRDRDRDRDRDRDRDRDRDRDRDRDRDR
YNSNSSSSPDDGKYKRDMDRDKYREKRFDSSRNNEHDRSSIRPLSDYGDYDDY
GRSKNKTEQYKEMVASNGNKRLEDSPIDLPENREAREPLRNAPTKGLHQPLGK
EVVMQCKKCFYHRAICTQTNPAEYIKLEKREDPMIYNDHDDFGFNNKNNIN
NNNNNNNFIHKDRINKNTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
EKKRKRKKKKIKASKKSHRHSDDSDSGSGSSGSSDLSKDESYSSNESE"
join(39790..40108,40252..47744)
/notes="ORF ID:dd_01572"
/codon_start=1
/product="Putative Acyltransferase"
/protein_id="AAL99334.1"

```

Query Match	Best Local Similarity	Score	DB 2	Length
Matches 72	Conservative 0	Mismatches 56	Indels 0	Gaps 0

```

QY 360 TATCTTCATCCATAGATTTCATAACTCTTGATCCATCTTCTATGTTTCAAGTG 419
Db 93202 TATTTGAATCAATTTTAAATTTATATTGCTTTTATTCATAATGTTGATTAATAA 93261

QY 420 TATAATTAGAGAGATCGATGATATATAATAAGTAAAGCTACCGTATCACCATGT 479
Db 93262 AATAATTGAAAAAATAAGATAATAATTAATAATAAATCTTCAATATACACATGT 93321

QY 480 GATGATTT 487
Db 93322 AATTTTT 93329

```

AL359194 159919 bp DNA linear PRI 14-MAR-2001
Human DNA sequence from clone Rp11-326D19 on chromosome 13, complete sequence.

AL359194
VERSION AL359194.7 GI:13273756
KEYWORDS HTG.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 159919)
BLAKEV, S.
Direct Submission
Submitted (13-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerquest@sanger.ac.uk
On Mar 11, 2001 this sequence version replaced gi:12963937.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; wp, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep. This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13

Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Ch13>
RP11-326D19 is from the library RPC1-11.2 constructed by the group
of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
VECTOR: pBACE3.6
This sequence is the entire insert of clone RP11-326D19. The true
right end of clone RP11-22917 is at 61034 in this sequence.

FEATURES

```
source
1..159919
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="13"
  /clone="RP11-326D19"
  /clone_1fb="RPC1-11.2"
  2..1157
    /note="L1PA5 repeat: matches 4987..6143 of consensus"
    repeat_region
    /note="17 copies 4 mer atag 88% conserved"
    repeat_region
    /note="30 copies 2 mer at 75% conserved"
    repeat_region
    /note="30 copies 2 mer at 75% conserved"
    repeat_region
    /note="21 copies 4 mer atata 71% conserved"
    repeat_region
    /note="35 copies 2 mer ta 75% conserved"
    repeat_region
    /note="7641"
    /note="L2 repeat: matches 2574..2638 of consensus"
    repeat_region
    /note="7996"
    /note="L2 repeat: matches 1323..1542 of consensus"
    repeat_region
    /note="8466"
    /note="L1PA5 repeat: matches 6035..6074 of consensus"
    repeat_region
    /note="9056..9128"
    /note="HAL1 repeat: matches 991..1062 of consensus"
    repeat_region
    /note="9847..10246"
    /note="L1P repeat: matches 1863..2261 of consensus"
    repeat_region
    /note="10247..10465"
    /note="L1PA5 repeat: matches 5920..6138 of consensus"
    repeat_region
    /note="11012..11318"
    /note="ALuB repeat: matches 1..298 of consensus"
    repeat_region
    /note="12421..12694"
    /note="L1P5 repeat: matches 5396..5669 of consensus"
    repeat_region
    /note="13079..13536"
    /note="L1MD2 repeat: matches 5638..6094 of consensus"
    repeat_region
    /note="13537..13829"
    /note="ALuS repeat: matches 5..300 of consensus"
    repeat_region
    /note="13830..13977"
    /note="L1MD2 repeat: matches 5491..5638 of consensus"
    repeat_region
    /note="14104..14291"
    /note="ALuB repeat: matches 93..282 of consensus"
    repeat_region
    /note="16400..16453"
    /note="L2 repeat: matches 2 mer ta 77% conserved"
    repeat_region
    /note="16425..16452"
    /note="7 copies 4 mer atata 100% conserved"
    repeat_region
    /note="16821..16856"
    /note="18 copies 2 mer ta 86% conserved"
    repeat_region
    /note="20660..21162"
    /note="L1MB8 repeat: matches 5649..6161 of consensus"
    repeat_region
    /note="21807..22019"
    /note="L1 repeat: matches 4045..4250 of consensus"
    repeat_region
    /note="22022..22105"
    /note="21 copies 4 mer gaag 77% conserved"
    repeat_region
    /note="22202..22888"
    /note="L1M4 repeat: matches 3360..4073 of consensus"
    repeat_region
    /note="22889..23252"
    /note="L1M4 repeat: matches 3360..4073 of consensus"
    repeat_region
    /note="23253..23302"
    /note="L1M4 repeat: matches 3310..3360 of consensus"
    repeat_region
    /note="23303..23693"
    /note="L1M4 repeat: matches 3310..3360 of consensus"
    repeat_region
    /note="23694..23996"
    /note="L1M4 repeat: matches 3074..3310 of consensus"
    repeat_region
    /note="24000..24037"
    /note="19 copies 2 mer ta 81% conserved"
    repeat_region
    24040..24808
    /note="L1PA3 repeat: matches 5378..6146 of consensus"
    repeat_region
    25498..26057
    /note="L1M1F repeat: matches 4..541 of consensus"
    repeat_region
    26649..26903
    /note="ALuS repeat: matches 1..255 of consensus"
    repeat_region
    27125..27152
    /note="7 copies 4 mer atag 92% conserved"
    repeat_region
    27432..27509
    /note="MAD1 repeat: matches 8..80 of consensus"
    repeat_region
    28572..28675
    /note="L2 repeat: matches 210..314 of consensus"
    repeat_region
    29376..31199
    /note="L1 repeat: matches 3165..4993 of consensus"
    repeat_region
    31271..31678
    /note="L1M1B repeat: matches 1..389 of consensus"
    repeat_region
    31702..31739
    /note="L19 copies 2 mer aa 78% conserved"
    repeat_region
    31773..31867
    /note="MER44 repeat: matches 1..96 of consensus"
    repeat_region
    31871..32131
    /note="MER44 repeat: matches 474..724 of consensus"
    repeat_region
    32405..32952
    /note="L1M1B repeat: matches 5495..6052 of consensus"
    repeat_region
    32954..33113
    /note="L2 repeat: matches 332..510 of consensus"
    repeat_region
    35044..35456
    /note="L1M1B repeat: matches 1..413 of consensus"
    repeat_region
    35578..35916
    /note="L1M1B repeat: matches 1..413 of consensus"
    repeat_region
    36028..36144
    /note="L1M1B repeat: matches 1..120 of consensus"
    repeat_region
    36974..37277
    /note="ALuB repeat: matches 1..297 of consensus"
    repeat_region
    37492..37591
    /note="L50 copies 2 mer gt 68% conserved"
    repeat_region
    37595..37800
    /note="L2 repeat: matches 2082..2310 of consensus"
    repeat_region
    38474..38589
    /note="L1M1B repeat: matches 20..141 of consensus"
    repeat_region
    39167..39460
    /note="ALuS repeat: matches 9..312 of consensus"
    repeat_region
    41933..42581
    /note="MER67C repeat: matches 1..709 of consensus"
    repeat_region
    43968..43995
    /note="7 copies 4 mer tgta 96% conserved"
    repeat_region
    44350..44381
    /note="16 copies 2 mer at 93% conserved"
    repeat_region
    44397..44493
    /note="L1M1B repeat: matches 2566..2674 of consensus"
    repeat_region
    44656..44823
    /note="L1M1B repeat: matches 51..224 of consensus"
    repeat_region
    44979..45156
    /note="L1M1B repeat: matches 3020..3198 of consensus"
    repeat_region
    45756..45854
    /note="L1M1B repeat: matches 3..102 of consensus"
    repeat_region
    45888..46113
    /note="L1M1B repeat: matches 1124..1351 of consensus"
    repeat_region
    46752..46979
    /note="L1M1B repeat: matches 1380..1617 of consensus"
    repeat_region
    46979..47348
    /note="L1M1B repeat: matches 2050..4496 of consensus"
    repeat_region
    49989..50162
    /note="L1M1B repeat: matches 5352..5530 of consensus"
    repeat_region
    50166..50607
    /note="L1M1B repeat: matches 4..458 of consensus"
    repeat_region
    50613..50815
    /note="L1M1B repeat: matches 2535..2748 of consensus"
    repeat_region
    51701..51876
    /note="L1M1B repeat: matches 129..302 of consensus"
    repeat_region
    51993..52011
    /note="L1M1B repeat: matches 4110..4231 of consensus"
    repeat_region
    52868..52899
```

```

repeat_region /note="16 copies 2 mer gt 91% conserved"
54804..54928
/note="L2 repeat: matches 1714..1838 of consensus"
55708..55771
/note="16 copies 4 mer tatg 78% conserved"
58435..58781
/note="LTR33 repeat: matches 48..398 of consensus"
61425..62137
/note="L1MA9 repeat: matches 5549..6270 of consensus"
62138..62198
/note="L1MA9 repeat: matches 5317..5375 of consensus"
62235..62544
/note="AluY repeat: matches 1..308 of consensus"
62545..62669
/note="FLAM_A repeat: matches 2..131 of consensus"
62687..63166
/note="L1MA9 repeat: matches 4826..5309 of consensus"

Query Match 7.8%; Score 38.4; DB 9; Length 159919;
Best Local Similarity 55.4%; Pred. No. 9.1; Mismatches 57; Indels 0; Gaps 0;
Matches 72; Conservative 1;

QY 360 TATCTTCATCCATGCTAGATTTCATATACTCTTGATCCATCTCTTATGTGTTTCAAGTG 419
DB 69212 TCTCTTAAGAGAGGTCCTTAATCTCTTAATCTCTGCAACTACTAATTTGTTTATTCATGTC 69271
QY 420 TATAATTTAGAGAGATGCGATGATATATAATTAATTAAGTAAAGCTACGGTATCACCATGT 479
DB 69272 TACAATTCAGTATTAGAGAACATATATAATTTAATATACATATATGTAATGTAATGTTTGT 69331
QY 480 GATGATTTT 489
DB 69332 GATTTTCTTT 69341

```

RESULT 14

AL672211/C

LOCUS AL672211 80828 bp DNA linear VRT 16-AUG-2002
 DEFINITION Zebrafish DNA sequence from clone BUSM1-6A2 contains a novel gene similar to OGRF, a novel OGRF pseudogene, two pseudogenes on a novel LINE-like transposon, a novel reverse transcriptase pseudogene, the fez gene for forebrain embryonic zinc-finger protein, part of a novel gene for a protein similar to Ca2+ dependent activator and four CpG islands, complete sequence.

ACCESSION

AL672211.7 GI:20338538

KEYWORDS HTG; Ca2+ dependent activator; CpG island; fez; forebrain; OGRF; opioid growth factor receptor; transposon; zinc-finger protein.

SOURCE

zebrafish

ORGANISM

Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.

REFERENCE

AUTHORS

JOURNAL

TITLE Direct Submission
 Submitted (16-AUG-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zface@sanger.ac.uk
 Clone requests: clonerequest@sanger.ac.uk

COMMENT

On Apr 29, 2002 this sequence version replaced gi:20268901.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following

abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Emi, EMBL; Swi, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep BUSM1-6A2 is from a Zebrafish PAC library
 VECTOR: pCYPAC-6

----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: zface@sanger.ac.uk

Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhirong Bao and Sean Eddy, submitted), and those beginning 'drr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see
 http://www.Projects/D_rerio/fishmask.shtml his clone was isolated from a PAC library (Incyte Genomics, Inc.) and provided by G. Levkowitz (Genentech Inc., San Francisco, USA). This sequence is the entire insert of clone BUSM1-6A2. The true left end of clone BUSM1-113 is at 44958 in this sequence.

FEATURES

source

Location/Qualifiers
 1..80828
 /organism="Danio rerio"
 /db_xref="taxon:7955"
 /clone="BUSM1-6A2"
 /clone_lib="BUSM1"
 76..87
 /note="2.0 copies 6 mer CATATA 24% conserved"
 repeat_region 129..153
 /note="2.8 copies 9 mer CATTATTAT 34% conserved"
 repeat_region 138..160
 /note="7.7 copies 3 mer ATT 28% conserved"
 repeat_region 458..477
 /note="2.0 copies 10 mer AGTTTTCAT 40% conserved"
 repeat_region 541..557
 /note="4.2 copies 4 mer AAT 34% conserved"
 repeat_region 553..666
 /note="2.0 copies 7 mer TACAAAA 28% conserved"
 repeat_region 687..709
 /note="1.9 copies 12 mer TACTTAAATAC 46% conserved"
 repeat_region 702..717
 /note="2.0 copies 8 mer TTAATA 32% conserved"
 repeat_region 806..822
 /note="2.4 copies 7 mer ATAAAT 25% conserved"
 repeat_region 856..874
 /note="2.1 copies 9 mer TTGTGACTT 38% conserved"
 repeat_region 1084..1131
 /note="3.4 copies 14 mer TACATACATACATA 60% conserved"
 repeat_region 1088..1098
 /note="2.8 copies 4 mer CATG 22% conserved"
 repeat_region 1096..1107
 /note="2.0 copies 6 mer CATACA 24% conserved"
 repeat_region 1101..1133
 /note="8.2 copies 4 mer ACAT 66% conserved"
 repeat_region 1338..1405
 /note="Dr000340 repeat: matches 120..187 of consensus"
 repeat_region 1782..1799
 /note="18.0 copies 1 mer A 36% conserved"
 repeat_region 1815..1826
 /note="2.0 copies 6 mer GTCTCA 24% conserved"
 repeat_region 2405..2415
 /note="2.2 copies 5 mer AACAG 22% conserved"
 repeat_region 2552..2562
 /note="2.8 copies 4 mer TTG 22% conserved"
 repeat_region 2787..2950
 /note="4.1 copies 40 mer ACACACTGATACATATAAACTTGATGTCAGAACAC 149% conserved"
 repeat_region 2841..2852
 /note="2.0 copies 6 mer CATATA 24% conserved"
 repeat_region 2935..2946
 /note="2.0 copies 6 mer CATATA 24% conserved"

COMMENT

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Feb 16, 2002 this sequence version replaced gi:18250125.
 ----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc
 Contact: sapiens@wustl.wustl.edu
 ----- Summary Statistics

 Center project name: H_NH0182K03

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at http://www.chori.org
 VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-252K7, 2000 bp overlap. Actual start of this clone is at base position 106699 of RP11-252K7; actual end is at base position 126118 of RP11-182K3.

FEATURES

source	1. 126118 /organism="Homo sapiens" /db_xref="taxon:9606" /map="2" /clone="RP11-182K3" /clone_lib="RPCI-11" /rpt_family="A)n" /rpt_family="A)n" 2003..2229 /note="similar to Homo sapiens EST BF890821 (NID:gi2828280)"
repeat_region	4309..5520 /rpt_family="L2" 5605..5990 /rpt_family="MaLR"
misc_feature	6191..6700 /note="match to EST AL037617 (NID:gs928218)"
repeat_region	6596..7193 /rpt_family="L2"
misc_feature	7525..7604 /note="match to EST AL037954 (NID:gs407270)"
misc_feature	7525..7604 /note="match to EST BE811024 (NID:gi0243258)"
misc_feature	7525..7604 /note="match to EST BE811026 (NID:gi0243260)"
misc_feature	7525..7604 /note="match to EST BE926756 (NID:gi0452832)"

misc_feature	7525..7604 /note="match to EST BI859737 (NID:gi6000484)"
misc_feature	7525..7604 /note="similar to Bos taurus EST BI680079 (NID:gi5632987)"
misc_feature	7525..7604 /note="similar to Mus musculus EST BB615821 (NID:gi6456044)"
misc_feature	7536..7604 /note="match to EST AU134477 (NID:gi0995016)"
repeat_region	7606..7641 /rpt_family="AT_rich"
misc_feature	8473..8926 /note="match to EST BF935807 (NID:gi2353131)"
repeat_region	9078..9368 /rpt_family="Alu"
repeat_region	9379..9685 /rpt_family="Alu"
misc_feature	9778..9982 /note="match to EST AW816972 (NID:gi7909966)"
misc_feature	10105 /note="match to EST AL517947 (NID:gi2781440)"
misc_feature	10167 /note="match to EST BE811072 (NID:gi0243306)"
misc_feature	10204..10361 /note="match to EST AL037954 (NID:gs407270)"
misc_feature	10204..10361 /note="match to EST AU134477 (NID:gi0995016)"
misc_feature	10204..10361 /note="match to EST BE926756 (NID:gi0452832)"
misc_feature	10204..10361 /note="match to EST BI859737 (NID:gi6000484)"
misc_feature	10204..10361 /note="similar to Bos taurus EST BI680079 (NID:gi5632987)"
misc_feature	10204..10361 /note="similar to Mus musculus EST BB615821 (NID:gi6456044)"
misc_feature	10204..10342 /note="match to EST BE811024 (NID:gi0243258)"
misc_feature	10204..10325 /note="match to EST BE811026 (NID:gi0243260)"
misc_feature	10207..10361 /note="similar to Mus musculus EST BI100185 (NID:gi4551078)"
misc_feature	10213..10325 /note="match to EST BE811072 (NID:gi0243306)"
misc_feature	10223..10361 /note="match to EST AL517947 (NID:gi2781440)"
misc_feature	10319..10361 /note="similar to Bos taurus EST BE810028 (NID:gi0241140)"
repeat_region	11371..11466 /rpt_family="MER2_type"
repeat_region	11467..11735 /rpt_family="MER2_type"
repeat_region	11736..12026 /rpt_family="Alu"
repeat_region	12027..12481 /rpt_family="MER2_type"
misc_feature	12617..12774 /note="match to EST BG983142 (NID:gi4385864)"
repeat_region	12733..13000 /rpt_family="Alu"
repeat_region	13003..13026 /rpt_family="(TG)n"
repeat_region	13162..13187 /rpt_family="AT_rich"
misc_feature	13615..13994 /note="match to EST AA328955 (NID:gi1981442)"
misc_feature	13679..14248 /note="match to EST AI950302 (NID:gs742612) wp08502.x1"
misc_feature	13841..14227 /note="match to EST R93169 (NID:gi967335) yt69c02.s1"
misc_feature	13884..14235 /note="match to EST AA682675 (NID:gs2669956) zj20h11.s1"

```

repeat_region 14236..14256
misc_feature   /rpt_family="(A)n"
1438..14799
/nc="match to EST R82015 (NID:g858618) y106f08.r1"
repeat_region 15005..15305
misc_feature   /rpt_family="Alu"
17481..17958
/nc="match to EST BF771506 (NID:g12119406)"
repeat_region 17786..18083
misc_feature   /rpt_family="Alu"
18084..18129
repeat_region  /rpt_family="(TCTA)n"
18691..18712
misc_feature   /rpt_family="AT-rich"
19094..19636
repeat_region  /nc="match to EST AL037616 (NID:g5928217)"
19686..19709
repeat_region  /rpt_family="AT-rich"
19804..19825

```

Query Match 7.7%; Score 38.2; DB 9; Length 126118;

Best Local Similarity 53.7%; Pred. No. 10;

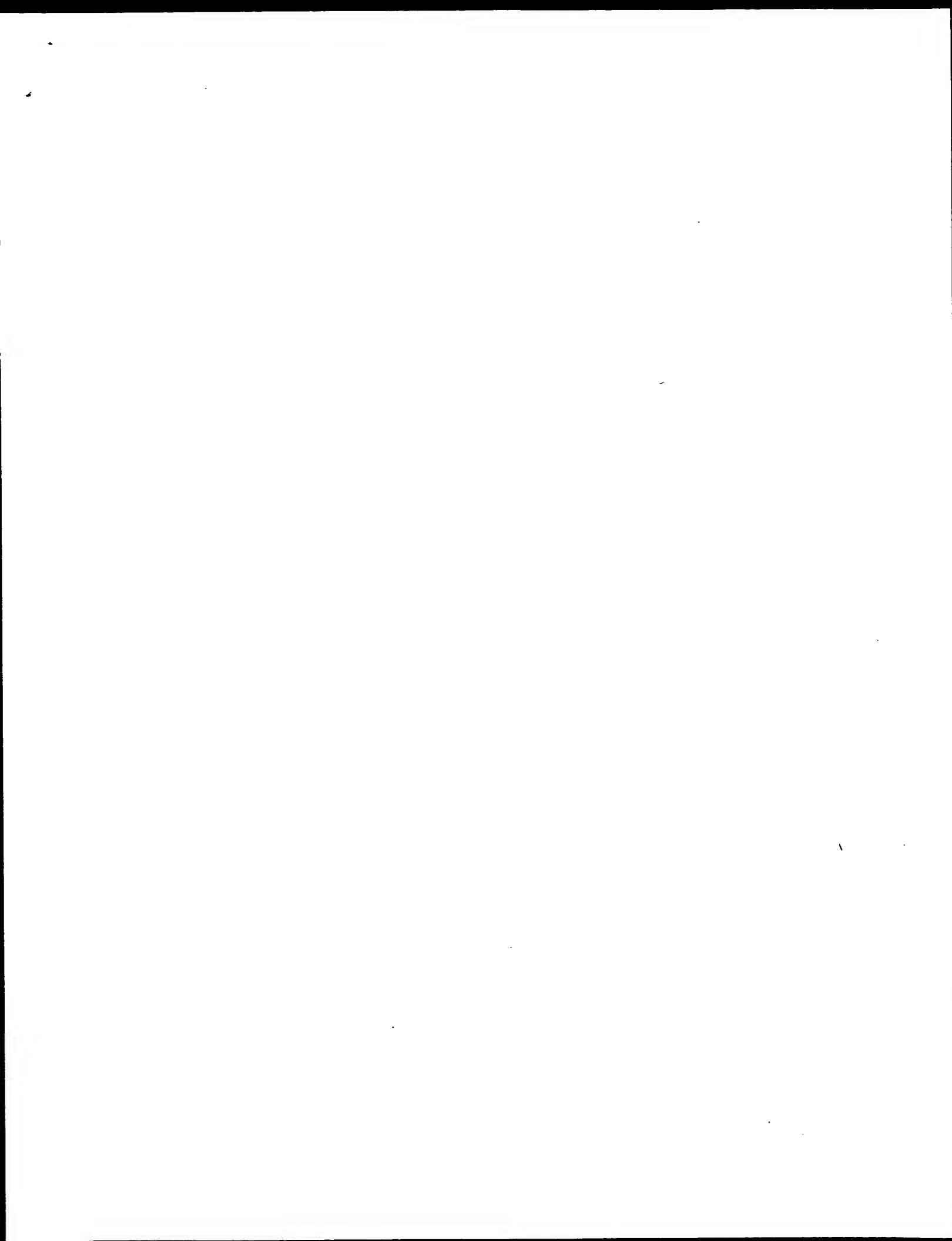
Matches 79; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

```

QY 331 AGGCGATGCAACCCCTTTGGTTGAGAGATATCTTCATCCATGCTAGATTTCATTAATC 390
Db 59680 AGTCCCTACCCGCTTGATCATCATGATAAGTACATCATGTGCTGACTGATGCTAGTT 59739
QY 391 TTGATCATCTTCTATGTTTTCAGTGTATTAATTAGAGATGATGATATATATA 450
Db 59740 TTGGTGCTTCTGCTGACTTTATATCTATTTATGAGCAGTGAATAGCTGATTAAT 59799
QY 451 AATAGTAAAGTACGATACCAT 477
Db 59800 ATGAATCAATGTGAGCAAAAAAAT 59826

```

Search completed: January 12, 2003, 08:14:38
Job time : 3376 secs



GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 12, 2003, 07:06:09 ; Search time 253 Seconds

(without alignments)
4388.281 Million cell updates/sec

Title: US-09-882-434A-2

Perfect score: 493

Sequence: 1 attagctcttgagctccat.....ccatgctgattcttacc 493

Scoring table:

IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

1: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1980.DAT.*
2: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1981.DAT.*
3: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1982.DAT.*
4: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1983.DAT.*
5: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1984.DAT.*
6: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1985.DAT.*
7: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1986.DAT.*
8: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1987.DAT.*
9: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1988.DAT.*
10: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1989.DAT.*
11: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1990.DAT.*
12: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1991.DAT.*
13: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1992.DAT.*
14: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1993.DAT.*
15: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1994.DAT.*
16: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1995.DAT.*
17: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1996.DAT.*
18: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1997.DAT.*
19: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1998.DAT.*
20: /SID82/gcgdata/geneeq/geneeqn-emb1/NA1999.DAT.*
21: /SID82/gcgdata/geneeq/geneeqn-emb1/NA2000.DAT.*
22: /SID82/gcgdata/geneeq/geneeqn-emb1/NA2001A.DAT.*
23: /SID82/gcgdata/geneeq/geneeqn-emb1/NA2001B.DAT.*
24: /SID82/gcgdata/geneeq/geneeqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	480	97.4	480	18	AAT88851
2	35.8	7.3	4667	23	ABL15860
3	35.8	7.3	5823	23	ABL15858
4	35.6	7.2	8011	24	ABL32078
5	35.6	7.2	8011	24	AA28368
6	35.4	7.2	8208	22	AAK84302
7	35.4	7.2	8209	22	AAK84300
8	35.4	7.2	8210	22	AAK84301
9	35.2	7.1	6375	24	ABL34024

10	35.2	7.1	38342	22	AA56745	Tumour suppressor
11	35.2	7.1	38342	22	ABK31506	Signal transductio
12	34.8	7.1	179	21	AA36785	Human dyserletin re
13	34.8	7.1	179	21	AA82876	Human dyserletin DN
14	34.4	7.0	374	20	AA877102	EST clone BH281.
15	34	6.9	7978	23	ABL03306	Drosophila melanog
16	33.4	6.8	777	24	ABK83174	Iron uptake ABC tr
17	33.4	6.8	9223	19	AAV52192	Streptococcus pneu
18	33.4	6.8	25577	22	AAK65015	Human immune/haema
19	33.4	6.8	25577	22	AAK65015	Human immune/haema
20	33.4	6.8	27540	23	AAK79382	Pneumococcal patho
21	33.2	6.7	4936	24	ABK03161	Drosophila melanog
22	33.2	6.7	139389	22	ABK03286	Human CDNA differe
23	33	6.7	8846	22	ABK64795	Human CDNA differe
24	33	6.7	62909	22	ABA15292	Human nervous syst
25	32.8	6.7	402	22	AAK28545	Genomic fragment #
26	32.8	6.7	5497	24	AA184087	Human polynucleoti
27	32.8	6.7	11220	21	AAK48974	Human immune syste
28	32.8	6.7	13784	24	AAK40061	Arabidopsis thalia
29	32.6	6.6	6041	24	AAK28377	Human chemically p
30	32.6	6.6	17142	22	AAK45498	Human chemically p
31	32.6	6.6	17142	22	ABK34107	Human immune syste
32	32.6	6.6	17142	24	ABK28430	DNA transcription
33	32.4	6.6	463	22	ABA59299	Human foetal liver
34	32.4	6.6	463	22	ABA28022	Probe #6488 for ge
35	32.4	6.6	463	22	AAK07518	Human brain expres
36	32.4	6.6	463	22	AAK33310	Human bone marrow
37	32.4	6.6	463	22	AA116407	Probe #6340 for ge
38	32.4	6.6	463	22	AA139099	Probe #785 used t
39	32.4	6.6	463	22	ABK08145	Human genome-deriv
40	32.4	6.6	9789	17	AAK41852	CDNA encoding plas
41	32.4	6.6	201143	24	ABK83568	Human DNA differen
42	32.2	6.5	397	22	AA183912	Human polynucleoti
43	32.2	6.5	1422	24	ABN92473	Staphylococcus epi
44	32.2	6.5	8440	24	ABK28346	DNA transcription
45	32.2	6.5	11046	24	AA516913	Human G-protein co

ALIGNMENTS

RESULT 1	
AAT88851	
ID	AAT88851 standard; cDNA; 480 BP.
XX	
AC	AAT88851;
XX	
DT	27-MAR-1998 (first entry)
XX	
DE	CDNA for Macadamia integrifolia antimicrobial protein 1.
XX	
KW	Antimicrobial protein 1; growth inhibition; fungus; bacterium;
KW	fungi; bacteria; pathogen control; ss.
XX	
OS	Macadamia integrifolia.
XX	
FH	Key
FT	Location/Qualifiers
FT	CD5
FT	70..378
FT	/*tag= a
FT	sig_peptide
FT	70..147
FT	/*tag= b
FT	mat_peptide
FT	148..375
FT	/*tag= c
FT	/product= antimicrobial_protein_1
XX	
XX	
XX	W09728185-A1.
XX	
XX	07-AUG-1997.
XX	
XX	31-JAN-1997;
XX	97WO-AU00052.
XX	
XX	31-JAN-1996;
XX	96AU-0007802.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Db 2563 TATTGTAGATGCACTCACTATTAAATTTTTCGATTTCGACATTTTTC 2622
 Qy 405 TATGTTTTCAAGTGTATTAATTAAGAGATGATATATAATTAAGTAAAGCT 464
 Db 2623 TCACCTGTTCTTTCGAGATATATGAATGATTCATTTTAAATATAAGCGATTCT 2682
 Qy 465 ACGGT 469
 Db 2683 ACTGT 2687

RESULT 8

AAK4301
 ID AAK4301 standard; DNA, 8210 BP.

AC AAK4301;

DT 07-NOV-2001 (first entry)

DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:39113.

KM Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

XX Cytoelastic; gene therapy; vaccine; metastasis; ds.

OS Homo sapiens.

PN W0200157182-A2.

XX 09-AUG-2001.

PF 17-JAN-2001; 2001WO-US01354.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-022547.

PR 14-AUG-2000; 2000US-0225157.

PR 14-AUG-2000; 2000US-0225158.

PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226679.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226868.

PR 23-AUG-2000; 2000US-0227182.

PR 30-AUG-2000; 2000US-0227009.

PR 01-SEP-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.

PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.

PR 05-SEP-2000; 2000US-0229509.

PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.

PR 08-SEP-2000; 2000US-0230438.

PR 08-SEP-2000; 2000US-0231242.

PR 08-SEP-2000; 2000US-0231243.

PR 08-SEP-2000; 2000US-0231244.

PR 08-SEP-2000; 2000US-0231413.

PR 08-SEP-2000; 2000US-0231414.

PR 08-SEP-2000; 2000US-0233080.

PR 08-SEP-2000; 2000US-0233081.

PR 12-SEP-2000; 2000US-0231968.

PR 14-SEP-2000; 2000US-0232397.

PR 14-SEP-2000; 2000US-0233398.

PR 14-SEP-2000; 2000US-0233399.

PR 14-SEP-2000; 2000US-0233400.

PR 14-SEP-2000; 2000US-0233401.

PR 14-SEP-2000; 2000US-0233063.

PR 14-SEP-2000; 2000US-0233064.

PR 21-SEP-2000; 2000US-0234223.

PR 21-SEP-2000; 2000US-0234224.

PR 25-SEP-2000; 2000US-0234997.

PR 25-SEP-2000; 2000US-0234998.

PR 26-SEP-2000; 2000US-0235484.

PR 27-SEP-2000; 2000US-0235834.

PR 27-SEP-2000; 2000US-0235835.

PR 29-SEP-2000; 2000US-0236327.

PR 29-SEP-2000; 2000US-0236367.

PR 29-SEP-2000; 2000US-0236368.

PR 29-SEP-2000; 2000US-0236369.

PR 02-OCT-2000; 2000US-0236802.

PR 02-OCT-2000; 2000US-0237037.

PR 02-OCT-2000; 2000US-0237038.

PR 02-OCT-2000; 2000US-0237039.

PR 13-OCT-2000; 2000US-0237040.

PR 13-OCT-2000; 2000US-0239935.

PR 13-OCT-2000; 2000US-0239937.

PR 20-OCT-2000; 2000US-0240960.

PR 20-OCT-2000; 2000US-0241221.

PR 20-OCT-2000; 2000US-0241785.

PR 20-OCT-2000; 2000US-0241786.

PR 20-OCT-2000; 2000US-0241787.

PR 20-OCT-2000; 2000US-0241808.

PR 20-OCT-2000; 2000US-0241809.

PR 01-NOV-2000; 2000US-0241826.

PR 08-NOV-2000; 2000US-0244617.

PR 08-NOV-2000; 2000US-0246474.

PR 08-NOV-2000; 2000US-0246475.

PR 08-NOV-2000; 2000US-0246476.

PR 08-NOV-2000; 2000US-0246477.

PR 08-NOV-2000; 2000US-0246478.

PR 08-NOV-2000; 2000US-0246523.

PR 08-NOV-2000; 2000US-0246524.

PR 08-NOV-2000; 2000US-0246525.

PR 08-NOV-2000; 2000US-0246526.

PR 08-NOV-2000; 2000US-0246527.

PR 08-NOV-2000; 2000US-0246528.

PR 08-NOV-2000; 2000US-0246532.

PR 08-NOV-2000; 2000US-0246610.

PR 08-NOV-2000; 2000US-0246611.

PR 08-NOV-2000; 2000US-0246613.

PR 17-NOV-2000; 2000US-0249207.

PR 17-NOV-2000; 2000US-0249208.

PR 17-NOV-2000; 2000US-0249209.

PR 17-NOV-2000; 2000US-0249210.

PR 17-NOV-2000; 2000US-0249211.

PR 17-NOV-2000; 2000US-0249212.

PR 17-NOV-2000; 2000US-0249213.

PR 17-NOV-2000; 2000US-0249214.

PR 17-NOV-2000; 2000US-0249215.

PR 17-NOV-2000; 2000US-0249216.

DE Tumour suppressor gene derived chemically modified sequence #469.
XX Human; tumour suppressor gene; oncogene; antitumour; cytosolic;
KM cancer; Cpg dinucleotide; single-nucleotide polymorphism; SNP;
XX cytosine methylation; ds.
OS Homo sapiens.
XX WO200168912-A2.
XX
XX
XX 20-SEP-2001.
PD
XX
XX 15-MAR-2001; 2001WO-EP02955.
PF
XX
XX 15-MAR-2000; 2000DE-1013847.
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIC-) EPIGENOMICS AG.
PA
XX
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2001-602752/68.
XX
XX
XX Fragments of chemically modified genes associated with tumour suppressor
PT genes and oncogenes, useful in designing primers and probes for
PT analysing diseases associated with cytosine methylation state e.g.
PT cancer -
XX
XX
XX Claim 1; SEQ ID No 469; 27bp; English.
XX
XX The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and
CC oncogenes having a sequence taken from 536 (actually 533 since
CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
CC (58) and sequences complementary to (58). The nucleic acid may be a
CC polypeptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
CC form part of a set of probes for detecting the cytosine methylation state
CC and/or single nucleotide polymorphisms and also to be used in an
CC array for analysing diseases associated with Cpg dinucleotides e.g.
CC cancers and tumours. The probes can also be used in a method for
CC ascertaining genetic and/or epigenetic parameters for the diagnosis
CC and/or therapy of existing diseases or the predisposition to specific
CC diseases, by analysing cytosine methylations. The parameters may be
CC compared to another set of genetic and/or epigenetic parameters, the
CC differences serving as basis for diagnosis and/or prognosis events which
CC are disadvantageous to patients. The present sequence is one of the
CC 533 genomic sequences derived from tumour suppressor genes and
CC oncogenes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 38342 BP; 1153 A; 472 C; 7565 G; 19152 T; 0 other;
Query Match 7.1%; Score 35.2; DB 22; Length 38342;
Best Local Similarity 55.8%; Pred. No. 6.8;
Matches 67; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
OY 346 TTTGGTTGGAAGATGATTCATCCAAATGCTAGATTTCATTAACCTTGGATCCATCTTCT 405
DB 2241 TTTTGTGTAAGATTTATTTTATTTATGCGGCTTTTATTTAGTTTATTTATGCTGA 2300
OY 406 ATGTTTTCAGTGTATTAATTAGAGATGATGATATTAATTAATAAGTAAAGCTA 465
DB 2301 TTATTAATAAATGTTTATTTATTTATGATTAATTTATTAATAAATTAAGTATTTAAAGATA 2360
RESULT 11

ABK31506
ID ABK31506 standard; DNA; 38342 BP.
XX
XX
XX AC ABK31506;
XX
XX 23-APR-2002 (first entry)
XX
XX
XX Signal transduction associated gene modified DNA #175.
DE
XX
XX
XX Human; signal transduction associated gene; cytosine methylation state;
KM Cpg island; signal transduction associated disease; solid tumour; cancer;
XX antitumour; cytosolic; mutant; ds.
XX
XX
XX Homo sapiens.
OS Synthetic.
OS
XX
XX WO200200926-A2.
PN
XX
XX 03-JAN-2002.
PD
XX
XX 29-JUN-2001; 2001WO-EP07472.
PF
XX
XX 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
XX
XX (EPIC-) EPIGENOMICS AG.
PA
XX
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-147896/19.
XX
XX
XX Oligonucleotide for diagnosis and therapy of diseases associated with
PT signal transduction e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with signal transduction -
XX
XX
XX Claim 1; SEQ ID No 349; 24bp; English.
XX
XX The present invention relates to chemically modified DNA sequences of
CC signal transduction associated genes. The DNA sequences are chemically
CC modified using a solution of bisulphite, hydrogen sulphite or
CC disulphite. Also disclosed are oligonucleotides and/or PNA oligomers
CC for detecting the cytosine methylation state (Cpg islands) of these
CC genes, and a method for the diagnosis and/or therapy of genetic and
CC epigenetic parameters of genes associated with signal transduction.
CC The genomic DNA can be obtained from cells or cellular components which
CC contain DNA, e.g. cell lines, biopsies, blood, sputum, stool, urine,
CC cerebral-spinal fluid, tissue embedded in paraffin such as tissue from
CC eyes, intestine, kidney, brain, heart, prostate, lung, breast or liver,
CC histologic object slides, and all their possible combinations. The
CC sequences of the invention are useful for the diagnosis and therapy of
CC diseases associated with signal transduction e.g. solid tumours and
CC cancer. ABK31158-ABK31545 represent chemically pretreated genomic DNA
CC sequences of different genes associated with signal transduction, or
CC their complementary sequences.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
XX
SQ Sequence 38342 BP; 11153 A; 472 C; 7565 G; 19152 T; 0 other;
Query Match 7.1%; Score 35.2; DB 24; Length 38342;
Best Local Similarity 55.8%; Pred. No. 6.8;
Matches 67; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
OY 346 TTTGGTTGGAAGATGATTCATCCAAATGCTAGATTTCATTAACCTTGGATCCATCTTCT 405
DB 2241 TTTTGTGTAAGATTTATTTTATTTATGCGGCTTTTATTTAGTTTATTTATGCTGA 2300
OY 406 ATGTTTTCAGTGTATTAATTAGAGATGATGATATTAATTAATAAGTAAAGCTA 465
DB 2301 TTATTAATAAATGTTTATTTATTTATGATTAATTTATTAATAAATTAAGTATTTAAAGATA 2360

RESULT 12

AAA36785/C

ID AAA36785 standard; DNA; 179 BP.

XX AC AAA36785;

XX DT 03-AUG-2000 (first entry)

XX DE Human dysferlin related nucleotide sequence SEQ ID NO:43.

XX DE Human; dysferlin; mutant; identification; chromosome 2p12-14;
KW detection; muscular dystrophy; diagnosis; hereditary muscular dystrophy;
KW miyoshi myopathy; limb girdle muscular dystrophy; ds.
XX Homo sapiens.
XX WO200011016-A1.
XX PD 02-MAR-2000.
XX PF 25-AUG-1999; 99WO-US19394.
XX PR 25-AUG-1998; 98US-0097930.
XX PA (GEO) GEN HOSPITAL CORP.
XX PA (UYPI-) UNIV PITTSBURGH.
XX PI Brown RH, Liu J, Hoffman E, Chou F;
XX WPI; 2000-246531/21.

XX PT Dysferlin polynucleotide, its mutant form useful for diagnosis and
PT treatment of hereditary muscular dystrophies e.g. miyoshi myopathy and
PT limb girdle muscular dystrophy -
XX Disclosure; Page 101; 136pp; English.
XX The present invention describes an isolated dysferlin DNA of 20-25
XX nucleotides in length, comprising a nucleotide sequence specifically
XX selected from nucleotides 911-913, 929-948, 1019-1038, 1392-1411,
XX 1424-1443, 1484-1503, 1499-1518, 1543-1565, 1715-1734, 1714-1759,
XX 2241-2260, 2864-2883, 2978-2997, 3057-3076, 3198-3217, 3252-3271,
XX 4356-4375, 4665-4684, 5015-5034, 5610-5629, 5726-5735, 6035-6054,
XX 6179-6198, 6243-6263 and 6529-6548 of the human dysferlin nucleotide
XX sequence given in AAA36744. Dysferlin nucleotide sequences containing
XX specific mutations can be used for diagnosing a patient, a foetus or
XX a pre-embryo at risk of developing a dysferlin associated disorder by
XX detecting mutations in the dysferlin gene in biological samples from
XX patients. Alternatively, the biological sample containing genomic DNA
XX can be incubated with a restriction enzyme, preferably BanII, BspI286I,
XX RsaI, RhaI, HaeIII, BspI286, NlaIV, NlaIII, BclI, AclI, BstEII, PstI,
XX HaeI, AluI, AclI, Tsp509I, SalI, HincII, TagI, HinfI, TfiI, SfaNI or
XX FokI and the presence or absence of a restriction enzyme site in the
XX sample is detected as an indication of the presence or absence of a
XX particular mutation in the sample. Dysferlin polynucleotides are useful
XX for treating hereditary muscular dystrophies such as miyoshi myopathy
XX (MM) and limb girdle muscular dystrophy-2B (LGMB-2B). MM and LGMB-2B
XX map to the human chromosome 2p12-14 region between the genetic markers
XX D2S292 and D2S286. The present sequence represents a nucleotide sequence
XX given in the sequence listing of the present specification, but which
XX does not appear to be mentioned further.

SQ Sequence 179 BP; 42 A; 34 C; 66 G; 37 T; 0 other;

Query Match 7.1%; Score 34.8; DB 21; Length 179;

Best Local Similarity 65.4%; Pred. No. 0.6;

Matches 51; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 22 CATACCTCTTCTCTCCACCATTATACACTTATCAGCTACCTCAGCCATGCTCCACC 81

DB 155 CCTGCTCCCTTGTCTCTTACCATTTTCCCGCAAGCTGACCTCCACAAAGGGGTCCACC 96

QY 82 AAGTTGTTCTTCTCAGTC 99

DB 95 AAGTTCTTCTTCTTACTC 78

RESULT 13

AAA82876/C

ID AAX82876 standard; DNA; 179 BP.

XX AC AAX82876;

XX DT 30-JUN-2000 (first entry).

XX DE Human dysferlin DNA #24.

XX DE Dysferlin; anti-dystrophic; gene therapy; muscular dystrophy; human;
KW skeletal muscle cell; hereditary; Miyoshi myopathy; diagnosis;
KW limb girdle muscular dystrophy-2B; brain-specific; ss.
XX Homo sapiens.
XX WO200011157-A1.
XX PD 02-MAR-2000.
XX PF 25-AUG-1999; 99WO-US19395.
XX PR 25-AUG-1998; 98US-0097927.
XX PA (GEO) GEN HOSPITAL CORP.
XX PI Brown RH, Liu J, Aoki M, Ho MF, Matsuda-Asada C;
XX WPI; 2000-237646/20.

XX PT Novel dysferlin genes and related proteins useful for diagnosis, risk
PT identification and treatment of hereditary muscular dystrophies and
PT other dysferlin related disorders -
XX Claim 11; Page 111; 146pp; English.
XX This invention describes a novel human dysferlin nucleic acid (I) and
XX its encoding protein (II), which has anti-dystrophic activity and can be
XX used for gene therapy. Introduction of (I), a vector comprising (I) or
XX dysferlin into a cell of a mammal can be used to decrease the symptoms of
XX muscular dystrophy. The dysferlin gene is normally expressed in skeletal
XX muscle cells and is selectively mutated in several families with the
XX hereditary muscular dystrophies, e.g. Miyoshi myopathy and limb girdle
XX muscular dystrophy-2B. The primers and oligonucleotides derived from (I)
XX can be used in diagnosis of or risk identification for dysferlin-related
XX disorders in patients, fetus, or pre-embryos. Expression of
XX brain-specific dysferlin may be important as a marker for normal neural
XX development. Dysferlin DNA or subgenomic coding sequences can be used for
XX therapy of the hereditary muscular dystrophies. This sequence represents
XX a fragment of the human dysferlin gene described in the method of the
XX invention.

SQ Sequence 179 BP; 42 A; 34 C; 66 G; 37 T; 0 other;

Query Match 7.1%; Score 34.8; DB 21; Length 179;

Best Local Similarity 65.4%; Pred. No. 0.6;

Matches 51; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 22 CATACCTCTTCTCTCCACCATTATCAGCTACCTCAGCCATGCTCCACC 81

DB 155 CCTGCTCCCTTGTCTCTTACCATTTTCCCGCAAGCTGACCTCCACAAAGGGGTCCACC 96

QY 82 AAGTTGTTCTTCTCAGTC 99

DB 95 AAGTTCTTCTTGTACTC 78

RESULT 14

AAV87102

ID AAV87102 standard; cDNA; 374 BP.
XX
AC AAV87102;
XX
DT 27-APR-1999 (first entry)
XX
DE EST clone BH281.
XX
KM Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolytic;
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
XX
OS Homo sapiens.
XX
PN MO9845435-A2.
XX
PD 15-OCT-1998.
XX
PF 10-APR-1998; 98MO-US06954.
XX
PR 10-APR-1997; 97US-0835913.
XX
PA (GEM) GENETICS INST INC.
XX
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI Racie LA, Spaulding V, Treacy M;
XX
DR WPI; 1999-070076/06.
XX
PT New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries
XX
PS Claim 1; Page 453; 633bp; English.
XX
CC This sequence represents an expressed sequence tag (EST), and is a
CC polynucleotide of the invention. The polynucleotides of the invention are
CC all secreted EST sequences isolated from a variety of human tissue
CC sources. The EST sequences and proteins encoded by them are predicted to
CC have useful biological activities which would make them suitable for
CC creating, preventing or ameliorating medical conditions in humans and
CC animals, although no supporting data is given. Suggested activities
CC include nutritional activity, immune stimulating or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The EST sequences are also stated to be useful for gene
CC therapy.
XX
SQ Sequence 374 BP; 132 A; 69 C; 48 G; 125 T; 0 other;

Query Match 7.0%; Score 34.4; DB 20; Length 374;
Best Local Similarity 55.2%; Pred. No. 1.2;
Matches 85; Conservative 1; Mismatches 67; Indels 1; Gaps 1;

QY 340 AACCTTTGGTGAAGATATCTTCATCCATGCTAGATTCATTAACCTTGATCA 399
DB 211 AATGCTTTTAAAGAAAAAACAATGAATCAATTAATAGTTT-TCACCTTAATTCATTT 269
QY 400 TCTTCTATGTTTTCAGATATTAATTAAGAGATGATGATATTAATTAATTAAGTAA 459
DB 270 TTGTTGTTGTTTAAATTTTCAATTAAGACAGATCAAGTATTAATTAAGAACTAAACG 329
QY 460 AAGTACGGTATGACCATGATGATTTTAAACC 493
DB 330 GAAGTACGAATTCCTCATTAATCACTTTCC 363

RESULT 15
ABL03306/c
ID ABL03306 standard; cDNA; 7978 BP.

XX
AC ABL03306;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 4400.
XX
KM Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li FWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
DR P-PSDB; ABB59203.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 4400; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB5737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 7978 BP; 2487 A; 1370 C; 1394 G; 2727 T; 0 other;

Query Match 6.9%; Score 34; DB 23; Length 7978;
Best Local Similarity 50.6%; Pred. No. 7.4;
Matches 82; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 329 CCAGGCGATGCAACCTTTTGGTGAAGATATCTTCATCCATGCTAGATTTCAATAC 388
DB 7302 CAAGGGGCACTCCGCTTTTGGTGAAGATATCTTCATCCATGCTAGATTTCAATAC 7243
QY 389 TCTTGATCATCTCTCTATGTTTTCAGATATTAATTAAGAGATGATGATATTA 448
DB 7242 ATTATCAATTCATTAATCACTTCATCACTTTATTAATTAATTAATTAATTA 7183
QY 449 TAAATAGTAAAGCTAGCATGATGATGATGATGATGATGATGATGATGATGATGAT 490
DB 7182 TACAGCTGAAATATATGCTTTATATACGAGCGATATTAATTTTA 7141

Search completed: January 12, 2003, 07:18:33
Job time : 284 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compen Ltd.

OW nucleic - nucleic search, using SW model

Run on: January 12, 2003, 07:12:54 ; Search time 49 Seconds
(without alignments)
3085.544 Million cell updates/sec

Title: US-09-882-434A-2

Sequence: 1 attaatctcttgatcctcat.....ccatgcatgatttacc 493

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 153338361 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: Issued_Patents_NA.*
2: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
3: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
5: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
6: /cgn2_6/ptodata/1/ina/Backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33.4	6.8	9223	US-08-961-527-59	Sequence 59, Appl
2	32.2	6.5	1422	US-09-134-001C-1936	Sequence 1936, Ap
3	31.8	6.5	537	US-08-845-539-5	Sequence 5, Appl1
4	31.6	6.4	537	US-09-362-642-5	Sequence 5, Appl1
5	31.6	6.4	6354	US-09-058-389A-5	Sequence 5, Appl1
6	31.6	6.4	6354	US-09-611-781-5	Sequence 5, Appl1
7	31.4	6.4	246240	US-08-724-394A-20	Sequence 20, Appl
8	31.4	6.4	246240	US-08-724-394A-21	Sequence 21, Appl
9	31.4	6.4	246240	US-08-724-394A-22	Sequence 22, Appl
10	31.2	6.3	4417	US-07-741-453A-57	Sequence 57, Appl
11	31.2	6.3	4301	US-08-121-446-3	Sequence 3, Appl1
12	31	6.3	112132	US-09-741-150-3	Sequence 3, Appl1
13	30.8	6.2	3715	US-09-600-776-9	Sequence 9, Appl1
14	30.6	6.2	5361	US-08-973-462-2	Sequence 2, Appl1
15	30.6	6.2	5361	US-08-973-462-1	Sequence 1, Appl1
16	30.4	6.2	2524	US-08-317-522A-1	Sequence 1, Appl1
17	30.4	6.2	2524	US-08-439-818A-1	Sequence 1, Appl1
18	30.4	6.2	2524	US-08-751-965-1	Sequence 1, Appl1
19	30.4	6.2	2524	US-08-738-975-1	Sequence 1, Appl1
20	30.4	6.2	2524	US-08-728-826-1	Sequence 1, Appl1
21	30.4	6.2	2524	US-08-808-599A-1	Sequence 1, Appl1
22	30	6.1	424	US-09-280-116-159	Sequence 159, App
23	30	6.1	450	US-08-090-523-28	Sequence 28, Appl
24	30	6.1	450	US-08-398-627-28	Sequence 28, Appl
25	30	6.1	450	US-08-406-857-2	Sequence 2, Appl
26	30	6.1	450	US-08-596-024-4	Sequence 4, Appl1
27	30	6.1	450	US-09-020-818-4	Sequence 4, Appl1

28	30	6.1	450	US-08-907-740-4	Sequence 4, Appl1
29	30	6.1	450	PCT-US94-07072-2	Sequence 2, Appl1
30	29.6	6.0	1517	US-08-963-602-6	Sequence 6, Appl1
31	29.6	6.0	1794	US-08-920-270-1	Sequence 1, Appl1
32	29.6	6.0	19124	US-08-487-826B-13	Sequence 13, Appl
33	29.4	6.0	634	US-09-328-111-825	Sequence 825, App
34	29.2	5.9	5852	US-07-867-106-2	Sequence 2, Appl1
35	29	5.9	1159	US-09-410-464-14	Sequence 14, Appl
36	28.8	5.8	618	US-08-684-024-3	Sequence 3, Appl1
37	28.8	5.8	618	US-09-145-868-3	Sequence 3, Appl1
38	28.8	5.8	973	US-09-177-650-119	Sequence 119, App
39	28.8	5.8	1037	US-08-240-124-1	Sequence 1, Appl1
40	28.8	5.8	1037	US-08-453-943-1	Sequence 1, Appl1
41	28.8	5.8	1037	US-09-057-121-1	Sequence 1, Appl1
42	28.8	5.8	1037	US-09-358-734-1	Sequence 1, Appl1
43	28.8	5.8	1070	US-08-299-567-8	Sequence 8, Appl1
44	28.8	5.8	1484	US-08-684-024-5	Sequence 5, Appl1
45	28.8	5.8	1484	US-09-145-868-5	Sequence 5, Appl1

ALIGNMENTS

RESULT 1
US-08-961-527-59/C
Sequence 59, Application US/08961527
Patent No. 6420135
GENERAL INFORMATION:
APPLICANT: Charles Kunesh
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 331
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 9223 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-59
Query Match 6.8%; Score 33.4; DB 4; Length 9223;
Best Local Similarity 52.6%; Pred. No. 1.2; Mismatches 62; Indels 0; Gaps 0;
Matches 70; Conservative 1;
Db 4971 TATTTTCAAAAATATATCATATATGTTCAATCTTCAGTAATTTTCAATATTTTCAGTGCAG 4912

GENERAL INFORMATION:
APPLICANT: Feder, John N.
APPLICANT: Kronmal, Gregory S.
APPLICANT: Lauer, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchihashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Fitts, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 017957-000100
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 246240 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..246240
OTHER INFORMATION: /note= "HLA-H.CONTIG"
US-08-724-394A-20

Query Match 6.4%; Score 31.4; DB 2; Length 246240;
Best Local Similarity 50.3%; Pred. No. 28; Mismatches 0; Gaps 0;
Matches 77; Conservative 0; Indels 76; Gaps 0;
QY 131 GTGAGATGGTGAATGGGAGTGCAATTTACAGTATGGAGTGGTCCAGGTTGTAAACAACCGTG 190
DB 145036 GGGAGAGGAGGACGGTGGTTATTTTATGCGCTCATAGGCTATATATTACACAATAGAGT 144977
QY 191 CTGAGCGATATAGCAAGTGTGGATGCTAGCTATACATCAGAGGAGGCTATGACTTCA 250
DB 144976 CATACATATTTAGCACGTTTGGGGGACAGCTATATATTATGAGGGGTGCCAAGTGCA 144917
QY 251 GCTACACTGGACAAACTGCTGCTCTCTACAACC 283
DB 144916 TTCACAATGGATAAACACGCTGTAATATACCTCC 144884

RESULT 8
US-08-724-394A-21/c
Sequence 21, Application US/08724394A
Patent No. 5872237
GENERAL INFORMATION:
APPLICANT: Feder, John N.
APPLICANT: Kronmal, Gregory S.
APPLICANT: Lauer, Peter M.
APPLICANT: Ruddy, David A.

APPLICANT: Thomas, Winston
APPLICANT: Tsuchihashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Fitts, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 017957-000100
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 246240 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..246240
OTHER INFORMATION: /note= "HLA-H.CONTIG"
US-08-724-394A-21

Query Match 6.4%; Score 31.4; DB 2; Length 246240;
Best Local Similarity 50.3%; Pred. No. 28; Mismatches 0; Gaps 0;
Matches 77; Conservative 0; Indels 76; Gaps 0;
QY 131 GTGAGATGGTGAATGGGAGTGCAATTTACAGTATGGAGTGGTCCAGGTTGTAAACAACCGTG 190
DB 145036 GGGAGAGGAGGACGGTGGTTATTTTATGCGCTCATAGGCTATATATTACACAATAGAGT 144977
QY 191 CTGAGCGATATAGCAAGTGTGGATGCTAGCTATACATCAGAGGAGGCTATGACTTCA 250
DB 144976 CATACATATTTAGCACGTTTGGGGGACAGCTATATATTATGAGGGGTGCCAAGTGCA 144917
QY 251 GCTACACTGGACAAACTGCTGCTCTCTACAACC 283
DB 144916 TTCACAATGGATAAACACGCTGTAATATACCTCC 144884

RESULT 9
US-08-724-394A-22/c
Sequence 22, Application US/08724394A
Patent No. 5872237
GENERAL INFORMATION:
APPLICANT: Feder, John N.
APPLICANT: Kronmal, Gregory S.
APPLICANT: Lauer, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchihashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Fitts, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 017957-000100
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 246240 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..246240
OTHER INFORMATION: /note= "HLA-H.CONTIG"
US-08-724-394A-21

NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND AND TOWNSEND AND CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Flets, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 017957-000100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 246240 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..246240
OTHER INFORMATION: /note= "HIV-H. CONTIG"
US-08-724-394A-22

Query Match
Best Local Similarity 50.3%; Score 31.4; DB 2; Length 246240;
Matches 77; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 131 GTGAGATGATGGAATGGAGTGCATTTTACAGTATGAGTGGTCCAGGTTGTAACACCGTG 190
DB 145036 GGGAGAGAGAGCAGCTGTTATTTTATGCTCATAGGCTATATTTACACATAGAGT 144977
QY 191 CTAGAGCATATAGCAAGTGTGATGCTCAGCTATACATCAGAGGAGGCTATGACTTCA 250
DB 144976 CATACATATTTAGCAGCTTTGGGGGAGCAGCTATATATTTATGAGGGGTCACAGTGCA 144917
QY 251 GCTACACTGACAACTGCTGCTCTCTACACC 283
DB 144916 TTCACATGATTAACACGCTGTAATATACCTCC 144884

RESULT 10
US-07-741-453A-57/C
Sequence 57, Application US/07741453A
Patent No. 6228597
GENERAL INFORMATION:
APPLICANT: PARMENTIER, MARC
APPLICANT: LIBERT, FREDERIC
APPLICANT: DUMONT, JACQUES
APPLICANT: VASSART, GILBERT
TITLE OF INVENTION: POLYPEPTIDES HAVING THYROTROPIN-RECEPTOR
TITLE OF INVENTION: ACTIVITY, NUCLEIC ACID SEQUENCES CODING FOR SUCH RECEPTORS
NUMBER OF SEQUENCES: 62
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DARBY & CUSHMAN
STREET: 1615 L STREET, N.W.
CITY: WASHINGTON, D.C.
COUNTRY: U.S.A.
ZIP: 20036

COMPUTER READABLE FORM:
MEDIUM TYPE: Disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/741,453A
FILING DATE: 19911015
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N.
REGISTRATION NUMBER: 16773
REFERENCE/DOCKET NUMBER: 91913/1107/US/ST
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 861-3000
TELEFAX: (202) 822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 4417 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-07-741-453A-57

Query Match
Best Local Similarity 45.8%; Score 31.2; DB 4; Length 4417;
Matches 108; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 85 TTGTTCTTCTCAGTCATTACTGTGTATGCTCATAGCAATGCGAATGATGATGAT 144
DB 499 TTGTCACATCAGAGAAATCTCAAGTCCAGTGTGAAATGCCAAGAACTTCAGAGCT 440
QY 145 GGGAGTGAATTAAGTATGAGTGTCCAGCTGTAAACCGTGTGACGATATAGC 204
DB 439 GGGAGCTCTTTTGGGCGTCAAGGCTTATGATGTTAAGCTTGTGATTCGGAATCTCT 380
QY 205 AAGTGTGATGCTCAGTATACATCAGAGGAGGCTATGACTTCAGCTACAGTACAGCA 264
DB 379 AAGTACATCTTTTAAATTTGAGAGAAATGTATTCAGCCGCTGCAAGATTGCA 320
QY 265 ACTGCTGCTCTTACACACAGCTGATGATGAGTGTGTCACACACAGGTTTGG 320
DB 319 TCTATTGACAAAGTATAGCTCGAAATATTGGCAGATTTGAAATGACAGACTGGG 264

RESULT 11
US-08-121-446-3/C
Sequence 3, Application US/08121446
Patent No. 6313276
GENERAL INFORMATION:
APPLICANT: IMURA, HIROO
APPLICANT: NAKAO, KAZUWA
APPLICANT: NAKANISHI, SHIGETADA
TITLE OF INVENTION: A HUMAN ENDOTHELIN RECEPTOR
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/121,446
FILING DATE:
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/911,684
; FILING DATE: 10-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: CIOTTI, THOMAS E.
; REGISTRATION NUMBER: 21,013
; REFERENCE/DOCKET NUMBER: 29900-20324.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 813-5600
; TELEFAX: (415) 494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4301 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 238..1566
US-08-121-446-3

Query Match 6.3%; Score 31; DB 4; Length 4301;
Best Local Similarity 51.0%; Pred. No. 4.7; Mismatches 70; Indels 0; Gaps 0;
Matches 73; Conservative 0;
QY 344 CTTTGGTTGGAAGAGTATCTTCATCCCAATGCTAGATTTCATACTCTTGGATCCATCTT 403
DB 3673 CATTTTITTTAAAGCCCACTCAATGCAATTTTATTATAAATAATGCAAGTATGCTTTT 3614
QY 404 CTATGTTTTCAAGTGTATAATTAGAGAGATGCATGGATATATAATAAATAGTAAAGC 463
DB 3613 GTGTGTAGTTTCAAAATTTGGCAAGATAATTAGAGAACATGATATCTGACATGGAGAGGA 3554
QY 464 TACGGTATCACCATGTGATTT 486
DB 3553 AAGGTATCAGGATGTAATAATT 3531

RESULT 12
US-09-741-150-3/c
; Sequence 3, Application US/09741150
; Patent No. 6436689
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CLO00968
; CURRENT APPLICATION NUMBER: US/09/741,150
; CURRENT FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 112132
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(112132)
; OTHER INFORMATION: n = A,T,C or G
US-09-741-150-3

Query Match 6.3%; Score 31; DB 4; Length 112132;
Best Local Similarity 56.3%; Pred. No. 26; Mismatches 58; Conservative 0; Indels 45; Gaps 0;
QY 371 AATGCTAGATTTTCATAACTCTTGGATCCATCTTCTATGTTTTTCAAGTGTATAATTAGAG 430
DB 106795 AATGTGAAAAATGTGCAGTTTTTCTTTTATTGATCAAAAAATTTTAAATATGATGATAGTT 106736
QY 431 AGATCGATGATATATAATAAATAAGTAAAGCTACGGTATCA 473

Db 106735 ACAGGCATAAAATATATATATATACTTTAAACCAAGGTTTTCA 106693

RESULT 13
US-09-600-776-9
; Sequence 9, Application US/09600776
; Patent No. 6326168
; GENERAL INFORMATION:
; APPLICANT: Yamanouchi Pharmaceutical Co., Ltd.
; TITLE OF INVENTION: A novel potassium channel protein
; FILE REFERENCE: Y9903-PCT
; CURRENT APPLICATION NUMBER: US/09/600,776
; CURRENT FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: JP P1998-011434
; PRIOR FILING DATE: 1998-01-23
; PRIOR APPLICATION NUMBER: JP P1998-346198
; PRIOR FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 3715
; TYPE: DNA
; ORGANISM: Rattus sp.
US-09-600-776-9

Query Match 6.2%; Score 30.8; DB 4; Length 3715;
Best Local Similarity 61.0%; Pred. No. 5.1; Mismatches 32; Indels 0; Gaps 0;
Matches 50; Conservative 0;
QY 233 AGGAGGCTATGACTTCAGCTACACTGGACAACTGCTCTCTTACAAACAGGCTGGAT 292
DB 2772 AGTGAGCGAAGAAACACACACACTGGCAAGCTACGCGCGGTGACGGAGCTGTCT 2831
QY 293 GCAGTGGTGTTCACACACCAG 314
DB 2832 GAACAGGTGCTGCAGATGCGAG 2853

RESULT 14
US-08-973-462-2/c
; Sequence 2, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 5361
; TYPE: DNA
; ORGANISM: P. falciparum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(5361)
US-08-973-462-2

Query Match 6.2%; Score 30.6; DB 4; Length 5361;
Best Local Similarity 52.0%; Pred. No. 7.2; Mismatches 66; Conservative 1; Indels 60; Gaps 0;
QY 364 TTCATCAATGCTAGATTTCATCAACTCTTGGATCCATCTTCTATGTTTTTCAAGTGTATA 423
DB 5244 TACTTCTTATACATTATATATTTTTTTCTTACTAAATTTTTTCTTCAATTTGTTGTAATA 5185
QY 424 ATTAGAGATGATGATGATATATAATAAATAAGTAAAGCTACGGTATCACCATGTGATG 483

Db 5184 ATGTTTAAATCTTAAAGATATCAAAATCAATGAATAAACGGTCTCTCTTATTTT 5125
 QY 484 ATTTTAA 490
 Db 5124 ATTTATA 5118

RESULT 15

US-08-973-462-1/c
 ; Sequence 1, Application US/08973462B
 ; Patent No. 6191270
 ; GENERAL INFORMATION:
 ; APPLICANT: DRUIHE, PIERRE
 ; APPLICANT: DAUBERES, PIERRE
 ; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
 ; FILE REFERENCE: 0660-0125-0 PCT
 ; CURRENT APPLICATION NUMBER: US/08/973,462B
 ; CURRENT FILING DATE: 1998-02-06
 ; EARLIER APPLICATION NUMBER: PCT/FR96/00894
 ; EARLIER FILING DATE: 1996-06-12
 ; EARLIER APPLICATION NUMBER: FR 95/07007
 ; EARLIER FILING DATE: 1995-06-13
 ; NUMBER OF SEQ ID NOS: 29
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 6152
 ; TYPE: DNA
 ; ORGANISM: P. falciparum
 US-08-973-462-1

Query Match 5.28; Score 30.6; DB 4; Length 6152;
 Best Local Similarity 52.08; Pred. No. 7.7;
 Matches 66; Conservative 1; Mismatches 60; Indels 0; Gaps 0;

QY 364 TTCATCCAAATGCTAGATTTCATTAACCTTGATCCATCTTCTATGTTTCAAGTGATA 423
 Db 5488 TACTTCTATTAATTAATTTTCTTACTTAATTTTCTTCAATTTGTTGTAATA 5429
 QY 424 ATTAGAGATGATGATATATATATAATTAAGCTACGATACCAATGTGATG 483
 Db 5428 ATGTTTAAATCTTAAAGATATCAAAATCAATGAATAAACGGTCTCTCTTATTTT 5369
 QY 484 ATTTTAA 490
 Db 5368 ATTTATA 5362

Search completed: January 12, 2003, 08:42:01
 Job time : 365 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using aw model

Run on: January 12, 2003, 08:14:45 ; Search time 58 Seconds

(without alignments)
3740.878 Million cell updates/sec

Title: US-09-882-434A-2

Perfect score: 493

Sequence: 1 attaagctttagagctccat.....ccatgatgatttacc 493

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 38906 seqs, 220051671 residues

Total number of hits satisfying chosen parameters: 778172

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*

2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*

3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*

4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*

5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*

6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*

7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*

8: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*

9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*

10: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*

11: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*

12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*

13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*

14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	492.6	99.9	493	10	US-09-882-434A-2
2	32.4	6.6	463	10	US-09-864-761-6488
3	32	6.5	202001	10	US-09-734-674-3
4	32	6.5	465237	10	US-09-933-267A-1
5	31.6	6.4	133	10	US-09-867-701-10265
6	31.4	6.4	224	10	US-09-960-352-9535
7	31.4	6.4	614	10	US-09-764-864-168
8	31.4	6.4	638	10	US-09-764-864-604
9	31.2	6.3	335913	9	US-09-754-853A-2
10	31.2	6.3	335913	9	US-09-754-853A-3
11	31	6.3	4301	10	US-09-931-157-2
12	30.8	6.2	493	10	US-09-882-434A-2
13	30.8	6.2	3715	9	US-09-965-830-9
14	30.6	6.2	453	9	US-09-929-230-3
15	30.6	6.2	2203	10	US-09-739-907-39
16	30.6	6.2	5361	9	US-09-742-096-2
17	30.6	6.2	6352	9	US-09-742-096-1
18	30.2	6.1	488	9	US-10-046-935-456
19	30.2	6.1	488	9	US-09-878-178-456

20	30.2	6.1	939	9	US-09-938-842A-2062	Sequence 2062, Ap
21	30.2	6.1	1873	10	US-09-925-301-238	Sequence 238, App
22	30.2	6.1	2456	10	US-09-822-830A-114	Sequence 114, App
23	30	6.1	450	10	US-09-797-467-4	Sequence 4, Appl1
24	30	6.1	2830	9	US-09-900-714A-1	Sequence 1, Appl1
25	29.8	6.0	275	10	US-09-923-876-5512	Sequence 5512, Ap
26	29.8	6.0	1143	10	US-09-818-564-1	Sequence 1, Appl1
27	29.6	6.0	216	9	US-10-025-380-714	Sequence 714, App
28	29.6	6.0	216	10	US-09-922-217-714	Sequence 714, App
29	29.6	6.0	216	10	US-09-833-263-714	Sequence 714, App
30	29.6	6.0	262	10	US-09-878-574-11395	Sequence 11395, A
31	29.6	6.0	467	10	US-09-867-701-2045	Sequence 2045, A
32	29.6	6.0	2000	9	US-09-938-842A-3639	Sequence 3639, Ap
33	29.4	6.0	525	10	US-09-954-456-2038	Sequence 2038, Ap
34	29.4	6.0	402	10	US-09-919-580-735	Sequence 735, App
35	29.4	6.0	634	10	US-09-879-536-825	Sequence 825, App
36	29.4	6.0	1361	10	US-09-764-864-218	Sequence 218, App
37	29.4	6.0	1808	10	US-09-925-297-192	Sequence 192, App
38	29.4	6.0	10378	10	US-09-764-864-1938	Sequence 1938, Ap
39	29.4	6.0	13341	8	US-08-910-386A-1	Sequence 1, Appl1
40	29.2	5.9	202	10	US-09-969-373-380	Sequence 380, App
41	29.2	5.9	530	9	US-09-796-692-3253	Sequence 3253, App
42	29.2	5.9	1366	9	US-09-938-842A-3485	Sequence 3485, Ap
43	29.2	5.9	2000	9	US-09-938-842A-3486	Sequence 3486, Ap
44	29	5.9	185	10	US-09-878-574-11947	Sequence 11947, A
45	29	5.9	243	10	US-09-878-574-13309	Sequence 13309, A

ALIGNMENTS

RESULT 1

US-09-882-434A-2

Sequence 2, Appl1

Patent No. US20020108144A1

GENERAL INFORMATION: John M.

APPLICANT: Mannet, John M.

APPLICANT: Marcu, John Paul

APPLICANT: Gaultier, Kenneth C.

APPLICANT: Green, Jodie Lynn

TITLE OR INVENTION: ANTI-MICROBIAL PROTEIN

FILE REFERENCE: CULIN18.1CPICT1

CURRENT FILING DATE: 2001-06-15

PRIOR FILING DATE: 1999-07-30

PRIOR APPLICATION NUMBER: 09/364395

PRIOR FILING DATE: 1998-11-09

PRIOR APPLICATION NUMBER: PCT/AU97/00052

PRIOR FILING DATE: 1997-01-31

PRIOR APPLICATION NUMBER: AU PN 7802

PRIOR FILING DATE: 1996-01-31

NUMBER OF SEQ ID NOS: 21

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 2

LENGTH: 493

TYPE: DNA

ORGANISM: Macadamia integrifolia

FEATURE:

NAME/KEY: CDS

LOCATION: (70)...(375)

OTHER INFORMATION: Y=t or c.

US-09-882-434A-2

Query Match

Best Local Similarity 99.9%; Score 492.6; DB 10; Length 493;

Best Local Similarity 100.0%; Pred. No. 1.6e-141;

Matches 493; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATTAAGCTTtagagctccatctctccacacattagacttatcacta 60

DB 1 ATTAAGCTTtagagctccatctctctccacacattagacttatcacta 60

OY 61 ACCTGAGCATGCTTCCACCAAGTGTCTTCAGCTCATTTACTGTGATGATGCTATA 120

Db 61 ACCTCAGCATGGCTTCACCAAGTTGTTCTTCAGTCATTAAGTGTGCTCATTA 120
Qy 121 GCAATGGCAAGTGAAGTGGTGAATGGAGTGCATTTACAGTATGAGTGGTCCAGGTTGT 180
Db 121 GCAATGGCAAGTGAAGTGGTGAATGGAGTGCATTTACAGTATGAGTGGTCCAGGTTGT 180
Qy 181 AACAAACCGTGTGAGCGGATATAGCAAGTGTGGATGCTCAGTATACATCAGAAGGGAGGC 240
Db 181 AACAAACCGTGTGAGCGGATATAGCAAGTGTGGATGCTCAGTATACATCAGAAGGGAGGC 240
Qy 241 TATGACTTCAGTCACTACCTGGCAAACTGCTCTCTACAAACAGGCTGGATGCAAGTGT 300
Db 241 TATGACTTCAGTCACTACCTGGCAAACTGCTCTCTACAAACAGGCTGGATGCAAGTGT 300
Qy 301 GTTGCACACACAGGTTTGGGTCCAGTGCAGGGCATGCCAACCCCTTTTGGTTGGAAGAGT 360
Db 301 GTTGCACACACAGGTTTGGGTCCAGTGCAGGGCATGCCAACCCCTTTTGGTTGGAAGAGT 360
Qy 361 ATCTTCATCCAATGCTAGATTTCACTCTCTGGATCCATCTTCTATGTTTTCAGTGT 420
Db 361 ATCTTCATCCAATGCTAGATTTCACTCTCTGGATCCATCTTCTATGTTTTCAGTGT 420
Qy 421 ATAATTAGAGAGTGCATGGATATATATAAATAAGTAAAGCTACGGTATCACCATGTG 480
Db 421 ATAATTAGAGAGTGCATGGATATATATAAATAAGTAAAGCTACGGTATCACCATGTG 480
Qy 481 ATGATTTTACCC 493
Db 481 ATGATTTTACCC 493

RESULT 2

US-09-864-761-6488
; Sequence 6488, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR FILING DATE: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 6488
; LENGTH: 463
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC003025.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.5
US-09-864-761-6488
Query Match 6.6%; Score 32.4; DB 10; Length 463;
Best Local Similarity 60.0%; Pred. No. 2.4;
Matches 54; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
Qy 264 AACTGCTGCTCTCTACAAACAGGCTGGATGAGTGGTGTTCACACACAGGTTTGGTGC 323
Db 152 AGCTGCAGCTGTGTGCAGCCAGGAGAGGTTCTCTGAGCACATGCTAGGTGTCTGGC 211
Qy 324 CAGTCCAGGCGATCAACCCCTTTTGGTTG 353
Db 212 CAGTCCAGGCGAGGAGGATATAAATAAGTTG 241
RESULT 3
US-09-734-674-3/c
; Sequence 3, Application US/09734674
; Patent No. US20020081648A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Ming-Hui et al
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL001018
; CURRENT APPLICATION NUMBER: US/09/734,674
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 202001
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(202001)
; OTHER INFORMATION: n = A,T,C or G
US-09-734-674-3
Query Match 6.5%; Score 32; DB 10; Length 202001;
Best Local Similarity 56.7%; Pred. No. 77;
Matches 59; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
Qy 10 TTGAGTCTCATACATACCT 69
Db 150243 TTGAGTCTCATACATACCT 150184
Qy 70 ATGGCTTCCACCAAGTTGTCT 113

Db 150183 CTGAGTTTCATGATTCATTATATTAATTAATTAATTTAT 150140

RESULT 4

US-09-933-267A-1/c
Sequence 1, Application US/09933267A
Patent No. US20020123095A1

GENERAL INFORMATION:

APPLICANT: Kaluen, Francis et al.

TITLE OF INVENTION: Estrogen receptor alpha variants and

TITLE OF INVENTION: methods of detection thereof

FILE REFERENCE: CL000258C14

CURRENT APPLICATION NUMBER: US/09/933,267A

CURRENT FILING DATE: 2001-08-21

PRIOR APPLICATION NUMBER: 60/160626

PRIOR FILING DATE: 1999-10-20

PRIOR APPLICATION NUMBER: 60/183756

PRIOR FILING DATE: 2000-02-22

PRIOR APPLICATION NUMBER: 09/692414

PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: 09/768184

PRIOR FILING DATE: 2001-01-24

PRIOR APPLICATION NUMBER: 09/804076

PRIOR FILING DATE: 2001-03-13

PRIOR APPLICATION NUMBER: 09/826314

PRIOR FILING DATE: 2001-04-05

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 1

LENGTH: 465237

TYPE: DNA

ORGANISM: human

US-09-933-267A-1

Query Match

Best Local Similarity 58.3%; Score 32; DB 10; Length 465237;

Matches 56; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Qy 355 AAGATGATCTTCATCCATGCTGATTTCTATTAATCTTGATCCATCTTCTATGTTTTC 414

Db 307902 AATGGTGGCTCTAACAGAGCTGGTTCTGCTGCTGCTCTTATTTCTAGGCCGCAC 307843

Qy 415 AAGCTATATTAATGAGAGATGCATGATATATATA 450

Db 307842 ATGTTATATATCTGATGATGCAATATCATGTAGA 307807

RESULT 5

US-09-867-701-10265/c

Sequence 10265, Application US/09867701

Patent No. US20020132237A1

GENERAL INFORMATION:

APPLICANT: Aglate, Paul A.

APPLICANT: Jones, Robert

APPLICANT: Harlocker, Susan L.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER

FILE REFERENCE: 210121.497

CURRENT APPLICATION NUMBER: US/09/867,701

CURRENT FILING DATE: 2001-05-29

NUMBER OF SEQ ID NOS: 10912

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 10265

LENGTH: 133

TYPE: DNA

ORGANISM: Homo sapien

US-09-867-701-10265

Query Match

Best Local Similarity 6.4%; Score 31.6; DB 10; Length 133;

Matches 58; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Qy 360 TATCTTCATCCATGCTAGATTCATTAAGTCTTGATCCATCTTCTATGTTTTCAGTG 419

Db 108 TTTTGTAAAGCTGTATGTATTTATTAAGAAATGCAATTAATCTTCTATTAATTTTG 49

Qy 420 TATAATTAGAGATGCATGATATATATTAATTAAGTAA 461

Db 48 TAATTTAAATTAATTTGATCTATGAAATTTAAAAA 7

RESULT 6

US-09-960-352-9535

Sequence 9535, Application US/09960352

Patent No. US20020137139A1

GENERAL INFORMATION:

APPLICANT: Warren, Wesley C.

APPLICANT: Tao, Nengbing

APPLICANT: Bylet, John C.

APPLICANT: Mathialagan, Nagappan

TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION

FILE REFERENCE: 16511.006/37-21(10298)C

CURRENT APPLICATION NUMBER: US/09/960,352

CURRENT FILING DATE: 2001-09-24

NUMBER OF SEQ ID NOS: 15112

SEQ ID NO 9535

LENGTH: 224

TYPE: DNA

ORGANISM: Bos taurus

OTHER INFORMATION: Clone ID: 41-LIB3058-026-Q1-K1-C2

US-09-960-352-9535

Query Match

Best Local Similarity 6.4%; Score 31.4; DB 10; Length 224;

Matches 67; Conservative 0; Mismatches 41; Indels 1; Gaps 1;

Qy 380 TTTTATTAAGTCTTGATCCATCTTCTATGTTTTCAGTGTATTAAGAGAT-GCAT 438

Db 51 TTTTAAATTTTATTTATTTTCTATTTTTCATTTTCAATTTATTAATTTATTTATTCAT 110

Qy 439 GGATATATTAATTAATTAAGTAAAGCTAGCATGATGATGATTT 487

Db 111 TTTTATTAATTAATTAATTAAGTTTATTAATTAATTAATTT 159

RESULT 7

US-09-764-864-168/c

Sequence 168, Application US/09764864

Patent No. US20020132753A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PT223

CURRENT APPLICATION NUMBER: US/09/764,864

CURRENT FILING DATE: 2001-01-17

Prior application data removed - consult PALM or file wrapper

NUMBER OF SEQ ID NOS: 1792

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 168

LENGTH: 614

TYPE: DNA

ORGANISM: Homo sapiens

US-09-764-864-168

Query Match

Best Local Similarity 6.4%; Score 31.4; DB 10; Length 614;

Matches 50; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

Qy 228 TCAGAGGAGCTATGATCTTCACTGACATGACAACTGCTCTTCAACACGAGC 287

Db 353 TCAGAGGAGCTATGATCTTCACTGACATGACAACTGAGAGGTTTCTCTCAGCA 294

Qy 288 TGATGACAGTGTGTTGACA 308

```
Db 293 TGGGTTTTCTGGTCTTCAATA 273
US-09-764-864-604/c
; Sequence 604, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ23
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 604
; LENGTH: 638
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (15)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (43)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-764-864-604

Query Match 6.4%; Score 31.4; DB 10; Length 638;
Best Local Similarity 61.7%; Pred. No. 5.9;
Matches 50; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

Qy 228 TCAGAGGAGGCTATGACTTACAGTACACTGGACAACTGCTCTCTACACACGAGC 287
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 405 TCAGAGAGTTGGTACAACTTCTGCTGCACTGAACACACTGGTAGGTTCTCTCAGCA 346
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 288 TGGATGACAGTGGTGTTCACA 308
||| ||| ||||| |||||
Db 345 TGGGTTTTCTGGTCTTCAATA 325
||| ||| ||||| |||||

RESULT 9
US-09-754-853A-2/c
; Sequence 2, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
; APPLICANT: Hauge, Brian M.
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Parsons, Jeremy D.
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE REFERENCE: 38-10(15810)B
; CURRENT APPLICATION NUMBER: US/09/754,853A
; CURRENT FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174,880
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 2
; LENGTH: 335913
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (45163)..(45314), (45450)..(45509), (46941)..(48763), (48975)..(49573)
; OTHER INFORMATION: Clone ID: 240017_region_G3
US-09-754-853A-2

Query Match 6.3%; Score 31.2; DB 9; Length 335913;
Best Local Similarity 52.4%; Pred. No. 1.7e+02;
Matches 66; Conservative 1; Mismatches 59; Indels 0; Gaps 0;

Qy 364 TTCATCCCAATGCTAGATTTTCATACTCTTGATCCATCTTCTATGTTTTCACAGTGATA 423
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 185216 TTTTTCATAAGCTAGAAATTTCTATCTATTAGTTTCTTCAAAAAATTTATTTCTAACTTATG 185157
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 424 ATTAGAGAGATGCATGGATATATAATAAATAAGTAAAGCTACGGTATCACCATGTGATG 483
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 185156 AATAAACTAAATTTTAGTTTATATAAAAAAATTTTCAATTTTTTCTTACAAGCATTTATAA 185097
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 484 ATTTTY 489
|||||
Db 185096 ATTTT 185091
|||||

RESULT 10
US-09-754-853A-3/c
; Sequence 3, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
; APPLICANT: Hauge, Brian M.
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Parsons, Jeremy D.
; APPLICANT: Wang, Ming Li
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE REFERENCE: 38-10(15810)B
; CURRENT APPLICATION NUMBER: US/09/754,853A
; CURRENT FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174,880
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 3
; LENGTH: 335913
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (46798)..(48763), (48975)..(49573)
; OTHER INFORMATION: Clone ID: 240017_region_G3
US-09-754-853A-3

Query Match 6.3%; Score 31.2; DB 9; Length 335913;
Best Local Similarity 52.4%; Pred. No. 1.7e+02;
Matches 66; Conservative 1; Mismatches 59; Indels 0; Gaps 0;

Qy 364 TTCATCCCAATGCTAGATTTTCATACTCTTGATCCATCTTCTATGTTTTCACAGTGATA 423
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 185216 TTTTTCATAAGCTAGAAATTTCTATCTATTAGTTTCTTCAAAAAATTTATTTCTAACTTATG 185157
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 424 ATTAGAGAGATGCATGGATATATAATAAATAAGTAAAGCTACGGTATCACCATGTGATG 483
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 185156 AATAAACTAAATTTTAGTTTATATAAAAAAATTTTCAATTTTTTCTTACAAGCATTTATAA 185097
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 484 ATTTTY 489
|||||
Db 185096 ATTTT 185091
|||||

RESULT 11
US-09-931-157-2/c
; Sequence 2, Application US/099311157
; Patent No. US20020082414A1
; GENERAL INFORMATION:
; APPLICANT: Imura, Hiroo
; APPLICANT: Nakao, Kazuwa
; APPLICANT: Nakanishi, Shigetada
; TITLE OF INVENTION: Human Endothelin Receptor
; FILE REFERENCE: 299002032411
; CURRENT APPLICATION NUMBER: US/09/931,157
; CURRENT FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: 08/121,446
; PRIOR FILING DATE: 1993-09-14
; PRIOR APPLICATION NUMBER: 07/911,684
; PRIOR FILING DATE: 1992-07-10
```

PRIOR APPLICATION NUMBER: JP 3-172828
PRIOR FILING DATE: 1991-07-12
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 2
LENGTH: 4301
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (238) ... (1566)
US-09-931-157-2

Query Match
Best Local Similarity 6.3%; Score 31; DB 10; Length 4301;
Pred. No. 21;
Matches 73; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 344 CTTTGGTGAAGATGATCTTCATGCTAGTATTCATCTTGGATCCATCTT 403
DB 3673 CATTTTAAAAAGCAGCTGAAATGCAATTTTATTAATATGCAAGTATGCTTTT 3614
QY 404 CTATGTTTTCAGTGTATATTAAGAGATGATGATATATATTAATTAAGTAAAGC 463
DB 3613 GTGTAGTATTCATAATTTGGCAAGATATTAAGAAATGATATCTGATGGAAGGA 3554
QY 464 TACGATATCAGATGATGATTT 486
DB 3553 AAGGTATCAGATGATTAATAATT 3531

RESULT 12

US-09-882-434a-2/c
Sequence 2, Application US/09882434A
Patent No. US20020108144A1
GENERAL INFORMATION:
APPLICANT: Manners, John M.
APPLICANT: Marcus, John Paul
APPLICANT: Goulet, Kenneth C.
APPLICANT: Green, Jodie Lyn
TITLE OF INVENTION: ANTI-MICROBIAL PROTEIN
FILE REFERENCE: CULN18.1CP1C1
CURRENT APPLICATION NUMBER: US/09/882,434A
CURRENT FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: 09/364395
PRIOR FILING DATE: 1999-07-30
PRIOR APPLICATION NUMBER: 09/117615
PRIOR FILING DATE: 1998-11-09
PRIOR APPLICATION NUMBER: PCT/AU97/00052
PRIOR FILING DATE: 1997-01-31
PRIOR APPLICATION NUMBER: AU PN 7802
PRIOR FILING DATE: 1996-01-31
NUMBER OF SEQ ID NOS: 21
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 493
TYPE: DNA
ORGANISM: Macadamia integrifolia
FEATURE:
NAME/KEY: CDS
LOCATION: (70) ... (375)
OTHER INFORMATION: y=t or c.
US-09-882-434a-2

Query Match
Best Local Similarity 6.2%; Score 30.8; DB 10; Length 493;
Pred. No. 7.8;
Matches 74; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 219 AGGTATACATCAAGAGGAGGATGATCTTCACTACACTGAGCAAACTGCTCTCTTA 278
DB 364 AGATATCTTTCACCAAAAGGATGATGCTTGTGACATGAGCAAACTGCTGCTG 305
QY 279 CAACGAGCTGATGATGATGCTTGTGACACACAGGATTTGGTCCAGTCCAGGCGATG 338

DB 304 CAACACACTGATCCAGCTGTTGTAGAGAGCAGAGTTTGTCCAGTGTAGTGAAGT 245
QY 339 CAACCTTTTGGTGAAGATGATCTT 364
DB 244 CATAGCTCCCTTCTGATGTATAGCT 219

RESULT 13

US-09-965-830-9
Sequence 9, Application US/09965830
Patent No. US20020177201A1
GENERAL INFORMATION:
APPLICANT: Yamanouchi Pharmaceutical Co., Ltd.
TITLE OF INVENTION: A novel potassium channel protein
FILE REFERENCE: Y9903-PCT
CURRENT APPLICATION NUMBER: US/09/965,830
CURRENT FILING DATE: 2001-10-01
PRIOR APPLICATION NUMBER: 09/600,776
PRIOR FILING DATE: 2001-07-21
PRIOR APPLICATION NUMBER: JP P1998-346198
PRIOR FILING DATE: 1998-12-04
NUMBER OF SEQ ID NOS: 12
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 9
LENGTH: 3715
TYPE: DNA
ORGANISM: Rattus sp.
US-09-965-830-9

Query Match
Best Local Similarity 6.2%; Score 30.8; DB 9; Length 3715;
Pred. No. 23;
Matches 50; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 233 AGGAGGCTATGACTTCAGCTACACTGAGCAAACTGCTCTCTACACAGCTGAT 292
DB 2772 AGTGAAGCAAGAAACAGACACTGAGCAAGCTACGCGAGCGGTGACGAGCTGCT 2831
QY 293 GCAGTGTGTTGACACACACAG 314
DB 2832 GAACAGGTGCTGCTGATGCGAG 2853

RESULT 14

US-09-929-230-3/c
Sequence 3, Application US/09929230
Patent No. US20020161203A1
GENERAL INFORMATION:
APPLICANT: Sheppard, Paul O.
APPLICANT: Bishop, Paul D.
TITLE OF INVENTION: RATTLESNAKE VENOM GLAND PROTEINS
FILE REFERENCE: 00-72
CURRENT APPLICATION NUMBER: US/09/929,230
CURRENT FILING DATE: 2001-08-13
NUMBER OF SEQ ID NOS: 14
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 453
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: This degenerate nucleotide sequence encodes the
OTHER INFORMATION: amino acid sequence of SEQ ID NO:2.
NAME/KEY: misc feature
LOCATION: 9, 12, 18, 21, 24, 27, 30, 36, 39, 42, 45, 48, 51, 54, 57,
LOCATION: 66, 69, 78, 81, 105, 111, 117, 123, 135, 141, 153, 162,
LOCATION: 168, 171, 174, 177, 180, 189, 192, 204, 210, 213, 219, 222,
LOCATION: 225, 231, 234, 243, 246, 258, 261, 270, 297, 306, 312, 324
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc feature
LOCATION: 330, 333, 336, 339, 348, 351, 357, 360, 366, 372, 390, 393,
LOCATION: 405, 408, 411, 417, 432, 435, 438, 441, 447, 450, 453
OTHER INFORMATION: n = A,T,C or G
US-09-929-230-3

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 12, 2003, 07:11:59 ; Search time 1947 Seconds

(without alignments)
4100.860 Million cell updates/sec

Title: US-09-882-434a-2

Sequence: 1 attagctcttgagctcatc.....ccatgtgatgttacc 493

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:
1: em_esba:*
2: em_eschum:*
3: em_escin:*
4: em_escmu:*
5: em_esclov:*
6: em_escpl:*
7: em_escro:*
8: em_escro:*
9: gb_esc1:*
10: gb_esc2:*
11: gb_esc3:*
12: gb_esc4:*
13: gb_esc4:*
14: gb_esc5:*
15: em_escfun:*
16: em_escfun:*
17: gb_esc6:*
18: em_esc6:*
19: em_esc6:*
20: em_esc6:*
21: em_esc6:*
22: em_esc6:*
23: em_esc6:*
24: em_esc6:*
25: em_esc6:*
26: em_esc6:*
27: em_esc6:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	88.8	18.0	605	AW010330	ST04G06 P
2	88.8	18.0	726	BI416519	BI416519 haep001xd
3	86	17.4	750	BI416967	BI416967 haep002xp
4	85.6	17.4	598	AW043287	AW043287 ST1E09 P
5	85	17.2	698	BI416872	BI416872 haep002xj
6	84.4	17.1	531	BI416595	BI416595 haep001xm

C	7	84.4	17.1	633	13	BI416764	BI416764 haep002xd
C	8	84.4	17.1	639	13	BI416814	BI416814 haep002xy
C	9	84.4	17.1	721	13	BI416901	BI416901 haep002xj
C	10	81.8	16.6	530	13	BI416868	BI416868 haep002xj
C	11	81.8	16.6	697	13	BI416667	BI416667 haep001xm
C	12	81.2	16.5	531	13	BI416665	BI416665 haep001xm
C	13	75.8	15.4	374	13	BI416726	BI416726 haep002xa
C	14	57.4	11.6	476	13	BI416558	BI416558 haep001xf
C	15	57	11.6	294	10	BE662406	BE662406 ST62/ST62
C	16	52.6	10.7	216	13	BI416830	BI416830 haep002xh
C	17	48.6	9.9	825	13	BI416601	BI416601 haep001xi
C	18	45.2	9.2	815	13	BI416504	BI416504 haep001xc
C	19	42.6	8.6	620	12	BF299376	BF299376 067BDF09
C	20	41.6	8.4	717	13	BM160808	BM160808 EST563342
C	21	39.6	8.0	592	10	AM257904	AM257904 687063G11
C	22	39.4	8.0	455	13	BM160535	BM160535 EST563058
C	23	39.4	8.0	504	13	BI473509	BI473509 fdp3B10.Y
C	24	39.4	8.0	660	13	BM163034	BM163034 EST565557
C	25	39	7.9	769	17	AQ751030	AQ751030 HS-5576.B
C	26	38.8	7.9	455	17	BH721895	BH721895 BOHWT1JTR
C	27	38	7.7	576	17	AZ458762	AZ458762 IM0263F07
C	28	37.8	7.7	576	17	BE051991	BE051991 GA_Ea001
C	29	37.4	7.6	589	12	BG136361	BG136361 EST76803
C	30	37.2	7.5	358	13	BM359106	BM359106 CA_Ea001
C	31	37.2	7.5	977	17	CNS60631	AL407572 T7_end of
C	32	37	7.5	554	17	AZ012832	AZ012832 RPCI-23-3
C	33	37	7.5	558	12	BG733428	BG733428 347312 MA
C	34	36.8	7.5	539	17	AZ248820	AZ248820 RPCI-23-9
C	35	36.6	7.4	479	10	AM334443	AM334443 S34H9 ACS
C	36	36.6	7.4	996	9	AL568166	AL568166 AL568166
C	37	36.4	7.4	645	13	BI416777	BI416777 haep002xd
C	38	36.2	7.3	774	17	BH422732	BH422732 BOGRM52TF
C	39	36.2	7.3	795	17	BH444988	BH444988 BOGRM52TF
C	40	36	7.3	265	12	BE981479	BE981479 UI-M-CGP
C	41	36	7.3	265	12	BE981840	BE981840 UI-M-CGP
C	42	36	7.3	265	12	BE981879	BE981879 UI-M-CGP
C	43	36	7.3	265	12	BE981926	BE981926 UI-M-CGP
C	44	35.8	7.3	392	10	AW417434	AW417434 54088 MAR
C	45	35.8	7.3	412	14	C82886	C82886 C82886 rabb

ALIGNMENTS

RESULT 1
LOCUS AW010330 605 bp mRNA linear EST 10-SEP-1999
DEFINITION ST04G06 Pine Triplex shoot tip library pinus taeda cDNA clone
ACCESSION ST04G06, mRNA sequence.
VERSION AW010330.1 GI:5859108
KEYWORDS EST.
SOURCE loblolly pine.
ORGANISM Pinus taeda
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
1 (bases 1 to 605)
AUTHORS Whetten, R.W., Kinlaw, C.S., Retzel, E. and Seaderoff, R.R.
TITLE The Pine Gene Discovery Project
JOURNAL Unpublished (1999)
COMMENT Contact: Ross Whetten
Forest Biotechnology Group
North Carolina State University
Dept. of Forestry, NC State University, 6113 Jordan Hall, Raleigh, NC, 27695-8008
Tel: 919-515-7800
Fax: 919-515-7801
Email: rosewhetten@unity.ncsu.edu
Seq primer: 5' lambda Triplex2 Sequencing Primer.
Location/Qualifiers
1..605
/organism="Pinus taeda"
/db_xref="taxon:3352"

FEATURES

source

/clone="ST04C06"
 /clone_lib="Pine Triplex shoot tip library"
 /lab_host="E. coli BM25.8"
 /notes="Organ: shoot tips; Vector: Lambda Triplex; Site_1: SfiI (A); Site_2: SfiI (B); Shoot tips (approx. 2 cm from apex) were collected during the spring, frozen and used for mRNA isolation. The SMART-PCR method (Clontech) was used to prepare a library from 1 ug total RNA, using the Lambda Triplex vector. Plasmid subclones in pTriplex were recovered by cre-lox excision in E. coli strain BM25.8 and sequenced from the 5' end."

BASE COUNT 155 a 119 c 147 g 175 t 9 others
 ORIGIN

Query Match 18.0%; Score 88.8; DB 10; Length 605;
 Best Local Similarity 62.4%; Pred. No. 1.3e-14;
 Matches 176; Conservative 0; Mismatches 97; Indels 9; Gaps 2;

Qy 143 ATGGAGTGCAATTTACAGTATGGAGTTCAGGTTCTTAACAACCGTGTGAGCGATATA 202
 Db 118 AGGCGAGTATTTTCACTCGCGGCGAGGCGCTGTTGTAACAACCATGCTGCGGATACA 177
 Qy 203 GCAAGTGTGGATGCTCAGCTATACATCAGAAAG-----GGAGGCTATGACTTCAGCTACA 256
 Db 178 GCAATGTGGTGTCTTAATATCGCAACAACGTTTCATGGAGGATACGAGTTCATGTATC 237
 Qy 257 CTGGACAAACTGCTGCTCTTACAAACAGGCTGGATGAGTGGTGTGACACACACAGGT 316
 Db 238 AAGCCAGACCGCTGCGGCTTTACAAACGGACAACTGCAAGGGCGTTGCTCAAAACCGGT 297
 Qy 317 TTGGTCCAGTCCAG-----GCATGCAACCTTTTGGTTCGAAGAGTATCTTCATCCAAT 373
 Db 298 TTTCTAGCAGTGTAGTCAAGCTTTCAGCAGTGTGTTGGTGAAGAGTGTTCATCCAGT 357
 Qy 374 GCTAGATTTCAATCTCTTGATCCATCTTCTATGTTTCA 415
 Db 358 GCTGAGTCTGTACAACTTTATATTACTATATCACTATA 399

RESULT 2
 BI416519 726 bp mRNA linear EST 15-AUG-2001
 LOCUS
 DEFINITION hasp001xd12f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion annosum cDNA clone hasp001xd12f, mRNA sequence.

ACCESSION BI416519 GI:15187542
 VERSION BI416519
 KEYWORDS EST.
 SOURCE Pinus sylvestris/Heterobasidion annosum.
 ORGANISM Eukaryota; mixed EST libraries.
 REFERENCE 1 (bases 1 to 726)
 AUTHORS Asiegbu, F.O., Nahalkova, J., Choi, W., Stenlid, J. and Dean, R.A.
 TITLE Expressed sequence tags of randomly selected cDNA clones from the interaction of the root rot fungus (Heterobasidion annosum) with seedling roots of Scots pine (Pinus sylvestris)

JOURNAL Unpublished (2001)
 COMMENT Contact: Fred O. Asiegbu
 Dept. of Forest Mycology & Pathology
 Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden

Tel: +46 18 67 15 98
 Fax: +46 18 30 92 45
 Email: Fred.Asiegbu@mykopat.slu.se
 Seq primer: T7 primer.

FEATURES
 source

Location/Qualifiers
 1..726
 /organism="Pinus sylvestris/Heterobasidion annosum"
 /db_xref="taxon:169015"
 /clone="hasp001xd12f"
 /clone_lib="Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp)"
 /dev_stage="Seedling roots of scots pine were infected for

6 days with H. annosum"
 /note="Vector: pT-Adv; Site_1: EcoRI; The subtractive hybridization cDNA library was constructed from scots pine roots infected for 6-days with mycelia of Heterobasidion annosum (FPS)."

BASE COUNT 170 a 148 c 193 g 211 t 4 others
 ORIGIN

Query Match 18.0%; Score 88.8; DB 13; Length 726;
 Best Local Similarity 63.1%; Pred. No. 1.4e-14;
 Matches 173; Conservative 0; Mismatches 92; Indels 9; Gaps 2;

Qy 143 ATGGCAGTGCATTTTACAGTATGGAGTTCAGGTTGTAACAACCGTGTGAGCGATATA 202
 Db 197 AGGCGAGTATTTTCACTCGGTGGCAGGCGCTGTTGTAACAACCATGCTGCGGATACA 256
 Qy 203 GCAAGTGTGGATGCTCAGCTATACATCAGAAAG-----GGAGGCTATGACTTCAGCTACA 256
 Db 257 GCAATGTGGTGTCTTAATATCGGCAACAACGTTTCATGGAGGATACGAGTTCATGTATC 316
 Qy 257 CTGGACAAACTGCTGCTCTTACAAACAGGCTGGATGAGTGGTGTGACACACACAGGT 316
 Db 317 AAGGCGACAGTCTCTCTCTTACAAACCGCAACTGCAAGGGCGTTGCTCAGACCCGGT 376
 Qy 317 TTGGTCCAGTCCCA---GGGATGCAACCTTTTGGTTCGAAGAGTATCTTCATCCAAT 373
 Db 377 TTTCTAGCAGTGTATTAATCAAGCTTTCAGCAGTGTGTTGGTGAAGAGTGTTCATCCAAT 436
 Qy 374 GCTAGATTTCAATCTCTTGATCCATCTTCTAT 407
 Db 437 GCTGAGTCTGTGACAACTTTATATTACTATTAT 470

RESULT 3
 BI416967/c

LOCUS
 DEFINITION hasp002xpl5f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion annosum cDNA clone hasp002xpl5f, mRNA sequence.

ACCESSION BI416967
 VERSION BI416967
 KEYWORDS EST.

SOURCE Pinus sylvestris/Heterobasidion annosum.
 ORGANISM Eukaryota; mixed EST libraries.

REFERENCE 1 (bases 1 to 750)
 AUTHORS Asiegbu, F.O., Nahalkova, J., Choi, W., Stenlid, J. and Dean, R.A.
 TITLE Expressed sequence tags of randomly selected cDNA clones from the interaction of the root rot fungus (Heterobasidion annosum) with seedling roots of Scots pine (Pinus sylvestris)

JOURNAL Unpublished (2001)
 COMMENT Contact: Fred O. Asiegbu

Dept. of Forest Mycology & Pathology
 Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden

Tel: +46 18 67 15 98
 Fax: +46 18 30 92 45

Email: Fred.Asiegbu@mykopat.slu.se
 Seq primer: T7 primer.

FEATURES
 source

Location/Qualifiers
 1..750
 /organism="Pinus sylvestris/Heterobasidion annosum"
 /db_xref="taxon:169015"
 /clone="hasp002xpl5f"

/clone_lib="Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp)"
 /dev_stage="Seedling roots of scots pine were infected for 6 days with H. annosum"
 /note="Vector: pT-Adv; Site_1: EcoRI; The subtractive hybridization cDNA library was constructed from scots pine roots infected for 6-days with mycelia of Heterobasidion annosum (FPS)."

BASE COUNT 222 a 161 c 173 g 170 t 24 others

ORIGIN

Query Match 17.4%; Score 86; DB 13; Length 750;
 Best Local Similarity 65.6%; Pred. No. 9.2e-14;
 Matches 160; Conservative 0; Mismatches 75; Indels 9; Gaps 2;

143 ATGGAGTGCAATTACAGATGAGTGTCCAGTTGTAAACACCGTCTGACGATATA 202
 Db AGGGCAGTATTATTCACCTGCGGAGGCTGTGTAAACACGATGCGGATACA 182
 628 AGGGCAGTATTATTCACCTGCGGAGGCTGTGTAAACACGATGCGGATACA 569
 Db GCAAGTGTGATGCTCAGCTATACATCAGAG-----GGAGCTATGACTTCACTACA 256
 203 GCAAGTGTGATGCTCAGCTATACATCAGAG-----GGAGCTATGACTTCACTACA 256
 Db GCAAGTGTGATGCTCAGCTATACATCAGAG-----GGAGCTATGACTTCACTACA 509
 568 GCAAGTGTGATGCTCAGCTATACATCAGAG-----GGAGCTATGACTTCACTACA 316
 Db CTGACCAACTGCTGCTCTTACACACGCTGATGCTGTGTCACACACAGAGT 449
 257 CTGACCAACTGCTGCTCTTACACACGCTGATGCTGTGTCACACACAGAGT 449
 Db AAGGCCAGATGCTGCTCTTACACACGCTGATGCTGTGTCACACACAGAGT 373
 317 TTGGGTCCAGTCCCA---GGGATGCAACCTTTGTTGTAAGATATCTTCACTCAAT 373
 Db TTCTGCGCAGTATTATTCAGCTTGCAGCCCTTTGTTGTAAGATATCTTCACTCAAGT 389
 448 TTCTGCGCAGTATTATTCAGCTTGCAGCCCTTTGTTGTAAGATATCTTCACTCAAGT 389
 374 GCTA 377
 Db 388 GCTA 385

RESULT 4
 AM043287 598 bp mRNA linear EST 18-SEP-1999
 LOCUS ST31E09, Pine Triplex shoot tip library Pinus taeda cDNA clone

DEFINITION ST31E09, Pine Triplex shoot tip library Pinus taeda cDNA clone

ACCESSION AM043287
 VERSION AM043287.1 GI:5903816

KEYWORDS EST.
 SOURCE 10b101ly pine.
 ORGANISM Pinus taeda

REFERENCE 1 (bases 1 to 598) Retzel, E. and Sederoff, R.R.
 AUTHORS Whetten, R.W., Kinlaw, C.S., Retzel, E. and Sederoff, R.R.
 TITLE The Pine Gene Discovery Project
 JOURNAL Unpublished (1999)

COMMENT Contact: Ross Whetten
 Forest Biotechnology Group
 North Carolina State University
 Dept. of Forestry, NC State University, 6113 Jordan Hall, Raleigh
 NC, 27695-8008
 Tel: 919-515-7800
 Fax: 919-515-7801
 Email: rosswhetten@unity.ncsu.edu
 Seq primer: 5' lambda Triplex2 Sequencing Primer.

FEATURES
 Location/Qualifiers

1..598
 /organism="Pinus taeda"
 /db_xref="taxon:3352"
 /clone="ST31E09"
 /clone_lib="Pine Triplex shoot tip library"
 /lab_host="E. coli BM25.8"
 /note="Organ: shoot tips; Vector: lambda Triplex; Site 1: SfiI (A); Site 2: SfiI (B); Shoot tips (approx. 2 cm from apex) were collected during the spring, frozen and used for mRNA isolation. The SMART-PCR method (Clontech) was used to prepare a library from 1 ug total RNA, using the lambda Triplex vector. Plasmid subclones in pTriplex were recovered by cre-lox excision in E. coli strain BM25.8 and sequenced from the 5' end."

BASE COUNT 146 a 126 c 150 g 171 t 5 others

ORIGIN
 Query Match 17.4%; Score 85.6; DB 10; Length 598;
 Best Local Similarity 61.7%; Pred. No. 1.1e-13;
 Matches 174; Conservative 0; Mismatches 99; Indels 9; Gaps 2;

143 ATGGAGTGCAATTACAGATGAGTGTCCAGTTGTAAACACCGTCTGACGATATA 202

Db AGGGCAGTATTATTCACCTGCGGAGGCTGTGTAAACACGATGCGGATACA 182

203 GCAAGTGTGATGCTCAGCTATACATCAGAG-----GGAGCTATGACTTCACTACA 256

Db GCAAGTGTGATGCTCAGCTATACATCAGAG-----GGAGCTATGACTTCACTACA 242

183 GCAAGTGTGATGCTCAGCTATACATCAGAG-----GGAGCTATGACTTCACTACA 242

257 CTGACCAACTGCTGCTCTTACACACGCTGATGCTGTGTCACACACAGAGT 316

Db AAGGCCAGATGCTGCTCTTACACACGCTGATGCTGTGTCACACACAGAGT 302

317 TTGGGTCCAGTCCCA---GGGATGCAACCTTTGTTGTAAGATATCTTCACTCAAT 373

Db TTCTGCGCAGTATTATTCAGCTTGCAGCCCTTTGTTGTAAGATATCTTCACTCAAGT 362

374 GCTAATTTCAATCAGCTTGCATCTTCAATGTTTCA 415

Db 363 GCTAATTTCAATCAGCTTGCATCTTCAATGTTTCA 404

RESULT 5
 B1416872 698 bp mRNA linear EST 15-AUG-2001
 LOCUS B1416872
 DEFINITION hasp002xj10f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion annosum cDNA clone hasp002xj10f, mRNA sequence.

ACCESSION B1416872
 VERSION B1416872.1 GI:15187895
 KEYWORDS EST.
 SOURCE Pinus sylvestris/Heterobasidion annosum.
 ORGANISM Pinus sylvestris/Heterobasidion annosum

REFERENCE 1 (bases 1 to 698)
 AUTHORS Asiegbu, F.O., Nahalkova, J., Choi, W., Stenlid, J. and Dean, R.A.
 TITLE Expressed sequence tags of randomly selected cDNA clones from the interaction of the root rot fungus (Heterobasidion annosum) with seedling roots of Scots pine (Pinus sylvestris)

JOURNAL Unpublished (2001)
 COMMENT Contact: Fred O. Asiegbu
 Dept. of Forest Mycology & Pathology
 Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden
 Tel: +46 18 67 15 98
 Fax: +46 18 30 92 45
 Email: Fred.Asiegbu@mykopat.slu.se
 Seq primer: T7 primer.

FEATURES
 Location/Qualifiers

1..698
 /organism="Pinus sylvestris/Heterobasidion annosum"
 /db_xref="taxon:168015"
 /clone="hasp002xj10f"
 /clone_lib="Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp)"
 /dev_stage="Seedling roots of scots pine were infected for 6 days with H. annosum"
 /note="Vector: pT-Adv; Site 1: EcoRI; The subtractive hybridization cDNA library was constructed from scots pine roots infected for 6-days with mycelia of Heterobasidion annosum (FPS)."

BASE COUNT 174 a 143 c 173 g 208 t

ORIGIN
 Query Match 17.2%; Score 85; DB 13; Length 698;
 Best Local Similarity 65.4%; Pred. No. 1.7e-13;
 Matches 159; Conservative 0; Mismatches 75; Indels 9; Gaps 2;

143 ATGGAGTGCAATTACAGATGAGTGTCCAGTTGTAAACACCGTCTGACGATATA 202

Db 144 AGGGCAGTATTATTCACCTGCGGAGGCTGTGTAAACACGATGCGGATACA 203

Qy 203 GCAAGTGTGATGCTCAGCTATACATCAGAG-----GGAGCTATGACTTCACTACA 256

```

Db 204 GCAATGTGGGTGTTCTAATATCGGCAACAACGTTGATGAGGATACAGATTCTATGATC 263
QY 257 CTGGCAAACTGCTGCTCTCTACACACGAGGCTGGATGAGTGTGTTGCACACACAGGT 316
Db 264 AAGGCCAGACCGCTGCGGCTTACAACACGGCAACTGCAAGGGGTTGCTCAGACCCGGT 323
QY 317 TTGGGTCCAGTGCCA---GGGCATGCAACCCCTTTTGGTGGAGAGATATCTTCATCCAAT 373
Db 324 TTTCAGCAGTGTAAATCAAGCTTGCAGCAGTTTGGTTGGAAGAGTTTTTTCATCCAGT 383
QY 374 GCT 376
Db 384 GCT 386

RESULT 6
BI416595 531 bp mRNA linear EST 15-AUG-2001
LOCUS hasp001xh23f Heterobasidion annosum - Scots pine infection stage
DEFINITION subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion
annosum cDNA clone hasp001xh23f, mRNA sequence.
ACCESSION BI416595.1 GI:15187618
VERSION hasp001xh23f
KEYWORDS EST.
SOURCE Pinus sylvestris/Heterobasidion annosum.
ORGANISM Pinus sylvestris/Heterobasidion annosum
Eukaryota; mixed EST libraries.
REFERENCE 1 (bases 1 to 531)
AUTHORS Asiegbu, F.O., Nahalkova, J., Choi, W., Stenlid, J. and Dean, R.A.
TITLE Expressed sequence tags of randomly selected cDNA clones from the
interaction of the root rot fungus (Heterobasidion annosum) with
seedling roots of Scots pine (Pinus sylvestris)
JOURNAL Unpublished (2001)
COMMENT Contact: Fred O. Asiegbu
Dept. of Forest Mycology & Pathology
Swedish University of Agriculture, Box 7026, S-750 07, Uppsala,
Sweden
Tel: +46 18 67 15 98
Fax: +46 18 30 92 45
Email: Fred.Asiegbu@mykopat.slu.se
Seq primer: T7 primer.
FEATURES
    source
        1..531
            /organism="Pinus sylvestris/Heterobasidion annosum"
            /db_xref="taxon:169015"
            /clone="hasp001xh23f"
            /clone_lib="Heterobasidion annosum - Scots pine infection
            stage subtraction cDNA library (hasp)"
            /dev_stages="Seedling roots of scots pine were infected for
            6 days with H. annosum"
            /note="Vector: pT-Adv; Site 1: EcoRI; The subtractive
            hybridization cDNA library was constructed from scots
            pine roots infected for 6-days with mycelia of
            Heterobasidion annosum (FP5)."
BASE COUNT 115 a 117 c 138 g 161 t
ORIGIN
Query Match 17.1%; Score 84.4; DB 13; Length 531;
Best Local Similarity 65.2%; Pred. No. 2.4e-13;
Matches 159; Conservative 0; Mismatches 76; Indels 9; Gaps 2;

QY 143 ATGGAGTGCATTTACAGTATGGAGTGTCCAGGTGTGTAACACCGTGTGAGCGATATA 202
Db 141 AGGGCAGTTATTTCACTCGTGGGCGAGGCGCTGTTGTAACACCATGTGCTCGATACA 200
QY 203 GCAAGTGTGGATGTCTACATACATACAGAG-----GGAGGCTATGACTTCAGCTACA 256
Db 201 GCNAATGTGGGTGTTCTAATATCGGTACGACGTTTCATGAGGATACGAGTTCGTGTATC 260
QY 257 CTGGCAAACTGCTGCTCTCTACACACGAGGCTGGATGAGTGTGTTGCACACACAGGT 316
Db 261 AAGGCCAGACTGCTTCTGCTTACAACACGCGCACTGCAAGGGCTGTGCTCAGACCCGTT 320

```

```

QY 317 TTGGTCCAGTGCCA---GGGCATGCAACCCCTTTTGGTGGAGAGATATCTTCATCCAAT 373
Db 321 TTTCGCAAGTGTCTAATCAAGATTGCGAGCGGTTTGGTTGGAAGAGTTTCTTCATCCAGT 380
QY 374 GCTA 377
Db 381 GCTA 384

RESULT 7
BI416764/c 633 bp mRNA linear EST 15-AUG-2001
LOCUS hasp002xd04f Heterobasidion annosum - Scots pine infection stage
DEFINITION subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion
annosum cDNA clone hasp002xd04f, mRNA sequence.
ACCESSION BI416764
VERSION BI416764.1 GI:15187787
KEYWORDS EST.
SOURCE Pinus sylvestris/Heterobasidion annosum.
ORGANISM Pinus sylvestris/Heterobasidion annosum
Eukaryota; mixed EST libraries.
REFERENCE 1 (bases 1 to 633)
AUTHORS Asiegbu, F.O., Nahalkova, J., Choi, W., Stenlid, J. and Dean, R.A.
TITLE Expressed sequence tags of randomly selected cDNA clones from the
interaction of the root rot fungus (Heterobasidion annosum) with
seedling roots of Scots pine (Pinus sylvestris)
JOURNAL Unpublished (2001)
COMMENT Contact: Fred O. Asiegbu
Dept. of Forest Mycology & Pathology
Swedish University of Agriculture, Box 7026, S-750 07, Uppsala,
Sweden
Tel: +46 18 67 15 98
Fax: +46 18 30 92 45
Email: Fred.Asiegbu@mykopat.slu.se
Seq primer: T7 primer.
FEATURES
    source
        1..633
            /organism="Pinus sylvestris/Heterobasidion annosum"
            /db_xref="taxon:169015"
            /clone="hasp002xd04f"
            /clone_lib="Heterobasidion annosum - Scots pine infection
            stage subtraction cDNA library (hasp)"
            /dev_stages="Seedling roots of scots pine were infected for
            6 days with H. annosum"
            /note="Vector: pT-Adv; Site 1: EcoRI; The subtractive
            hybridization cDNA library was constructed from scots
            pine roots infected for 6-days with mycelia of
            Heterobasidion annosum (FP5)."
BASE COUNT 197 a 154 c 127 g 155 t
ORIGIN
Query Match 17.1%; Score 84.4; DB 13; Length 633;
Best Local Similarity 65.2%; Pred. No. 2.5e-13;
Matches 159; Conservative 0; Mismatches 76; Indels 9; Gaps 2;

QY 143 ATGGAGTGCATTTACAGTATGGAGTGTCCAGGTGTGTAACACCGTGTGAGCGATATA 202
Db 540 AGGGCAGTTATTTCACTCGTGGGCGAGGCGCTGTTGTAACACCATGTGCTCGGATACA 481
QY 203 GCAAGTGTGGATGTCTACATACATACAGAGG-----AGGCTATGACTTCAGCTACA 256
Db 480 GCNAATGTGGGTGTTCTAATATCGGCAACACGTTTCATGAGGATACGAGTTCGTGTATC 421
QY 257 CTGGCAAACTGCTGCTCTCTACACACGAGGCTGGATGAGTGTGTTGCACACACAGGT 316
Db 420 AAGGCCAGACTGCTTCTGCTTACAACACGCGCAACTGCAAGGGCTGTGCTCAGACCCGTT 361
QY 317 TTGGGTCCAGTGCCA---GGGCATGCAACCCCTTTTGGTGGAGAGATATCTTCATCCAAT 373
Db 360 TTTCGCAAGTGTCTAATCAAGCTTGCAGCAGTTTGGTTGGAAGAGTTTTTTCATCCAGT 301
QY 374 GCTA 377

```

DB 300 GCTA 297

LOCUS B1416814 639 bp mRNA linear EST 15-AUG-2001

DEFINITION haep002xj03f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (haep) Pinus sylvestris/Heterobasidion annosum cDNA clone haep002xj03f, mRNA sequence.

ACCESSION B1416814

VERSION B1416814

KEYWORDS B1416814.1 GI:15187837

SOURCE EST.

ORGANISM Pinus sylvestris/Heterobasidion annosum.

REFERENCE Pinus sylvestris/Heterobasidion annosum Eukaryota; mixed EST libraries.

AUTHORS 1 (bases 1 to 639)

TITLE Aiegbu, F.O., Nahlikova, J., Choi, W., Stenlid, J. and Dean, R.A. Expressed sequence tags of randomly selected cDNA clones from the interaction of the root rot fungus (Heterobasidion annosum) with seedling roots of Scots pine (Pinus sylvestris)

JOURNAL Unpublished (2001)

COMMENT Contact: Fred O. Aiegbu Dept. of Forest Mycology & Pathology Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden

FEATURES

source

1. .639

/organism="Pinus sylvestris/Heterobasidion annosum"

/db_xref="taxon:169015"

/clone="haep002xj03f"

/clone_lib="Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (haep)"

/dev_stage="Seedling roots of scots pine were infected for 6 days with H. annosum"

/note="Vector: PT-Adv; Site 1: EcoRI; The subtractive hybridization cDNA library was constructed from scots pine roots infected for 6-days with mycelia of Heterobasidion annosum (FPS)."

BASE COUNT 199 a 154 c 127 g 159 t

ORIGIN

Query Match 17.1%; Score 84.4; DB 13; Length 639;

Best Local Similarity 65.2%; Pred. No. 2.5e-13;

Matches 159; Conservative 0; Mismatches 76; Indels 9; Gaps 2;

DB 143 ATGGAGTGCATTTCAGATGAGTGTCCAGTTGTAAACCGTGTGAGGATATA 202

DB 547 AGGCGAGTATTTCACGTCGCGGCGAGGCGCTGTGTAACACAGCTCGCGGATACA 488

DB 203 GCAAGTGTGATGCTCAGCTATACATCAAGAAAGG-----AGCTATGACTTCACTACA 256

DB 487 GCAATGTGGGTGTTCTAATATCGGCAACAGTTTCATGCGAGTACGAGTTCGTATC 428

DB 257 CTGACAAACGTGCTCTCTACAAACGAGTGTGAGTGTGTCACACACAGGT 316

DB 427 AAGGCCAGACTGCTTCTGCTTACACAGCGCAACGTAAGAGGCGTGTGCTAGACCGGTT 368

DB 317 TTGGGTCCAGTGCCA---GGGCAATGCAACCTTTTGGTTGGAAGATATCTTCAAT 373

DB 367 TTTCTGGCAGTGTAAATCAAGCTTGCAGCGGTTTGGTTGGAAGATTTTTCATCCAGT 308

DB 374 GCTA 377

DB 307 GCTA 304

RESULT 9

B1416901

LOCUS B1416901 721 bp mRNA linear EST 15-AUG-2001

DEFINITION haep002xj03f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (haep) Pinus sylvestris/Heterobasidion annosum cDNA clone haep002xj03f, mRNA sequence.

ACCESSION B1416901

VERSION B1416901

KEYWORDS B1416901.1 GI:15187924

SOURCE EST.

ORGANISM Pinus sylvestris/Heterobasidion annosum.

REFERENCE Pinus sylvestris/Heterobasidion annosum Eukaryota; mixed EST libraries.

AUTHORS 1 (bases 1 to 721)

TITLE Aiegbu, F.O., Nahlikova, J., Choi, W., Stenlid, J. and Dean, R.A. Expressed sequence tags of randomly selected cDNA clones from the interaction of the root rot fungus (Heterobasidion annosum) with seedling roots of Scots pine (Pinus sylvestris)

JOURNAL Unpublished (2001)

COMMENT Contact: Fred O. Aiegbu Dept. of Forest Mycology & Pathology Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden

FEATURES

source

1. .721

/organism="Pinus sylvestris/Heterobasidion annosum"

/db_xref="taxon:169015"

/clone="haep002xj03f"

/clone_lib="Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (haep)"

/dev_stage="Seedling roots of scots pine were infected for 6 days with H. annosum"

/note="Vector: PT-Adv; Site 1: EcoRI; The subtractive hybridization cDNA library was constructed from scots pine roots infected for 6-days with mycelia of Heterobasidion annosum (FPS)."

BASE COUNT 165 a 146 c 177 g 227 t 6 others

ORIGIN

Query Match 17.1%; Score 84.4; DB 13; Length 721;

Best Local Similarity 65.2%; Pred. No. 2.6e-13;

Matches 159; Conservative 0; Mismatches 76; Indels 9; Gaps 2;

DB 143 ATGGAGTGCATTTCAGATGAGTGTCCAGTTGTAAACCGTGTGAGGATATA 202

DB 213 AGGCGAGTATTTCACGTCGCGGCGAGGCGCTGTGTAACACAGCTCGCGGATACA 272

DB 203 GCAAGTGTGATGCTCAGCTATACATCAAGAAAGG-----AGCTATGACTTCACTACA 256

DB 273 GCAATGTGGGTGTTCTAATATCGGCAACAGTTTCATGCGAGTACGAGTTCGTATC 332

DB 257 CTGACAAACGTGCTCTCTCTACAAACGAGTGTGAGTGTGTCACACACAGGT 316

DB 333 AAGGCCAGACTGCTTCTGCTTACACAGCGCAACGTAAGAGGCGTGTGCTAGACCGGTT 392

DB 317 TTGGGTCCAGTGCCA---GGGCAATGCAACCTTTTGGTTGGAAGATATCTTCAAT 373

DB 393 TTTCTGGCAGTGTAAATCAAGCTTGCAGCGGTTTGGTTGGAAGATTTTTCATCCAGT 452

DB 374 GCTA 377

DB 453 GCTA 456

RESULT 10

B1416868

LOCUS B1416868 530 bp mRNA linear EST 15-AUG-2001

DEFINITION haep002xj06f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (haep) Pinus sylvestris/Heterobasidion annosum cDNA clone haep002xj06f, mRNA sequence.

ACCESSION B1416868

VERSION B1416868.1 GI:15187891

KEYWORDS EST.
SOURCE Pinus sylvestris/Heterobasidion annosum.
ORGANISM Pinus sylvestris/Heterobasidion annosum
 Eukaryota; mixed EST libraries.
REFERENCE 1 (bases 1 to 530)
AUTHORS Asiegbu, F.O., Nahalkova, J., Choi, W., Stenlid, J. and Dean, R.A.
TITLE Expressed sequence tags of randomly selected cDNA clones from the interaction of the root rot fungus (Heterobasidion annosum) with seedling roots of Scots pine (Pinus sylvestris)
JOURNAL Unpublished (2001)
COMMENT Contact: Fred O. Asiegbu
 Dept. of Forest Mycology & Pathology
 Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden
 Tel: +46 18 67 15 98
 Fax: +46 18 30 92 45
 Email: Fred.Asiegbu@mykopat.slu.se
 Seq primer: T7 primer.
FEATURES Location/Qualifiers
 source
 1..530
 /organism="Pinus sylvestris/Heterobasidion annosum"
 /db_xref="taxon:169015"
 /clone="hasp002xj06f"
 /clone_lib="Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp)"
 /dev_stages="Seedling roots of scots pine were infected for 6 days with H. annosum"
 /note="Vector: pT-Adv; Site 1: EcoRI; The subtractive hybridization cDNA library was constructed from scots pine roots infected for 6-days with mycelia of Heterobasidion annosum (FP5)."
BASE COUNT 142 a 124 c 107 g 157 t
ORIGIN
 Query Match 16.6%; Score 81.8; DB 13; Length 530;
 Best Local Similarity 64.6%; Pred. No. 1.3e-12;
 Matches 157; Conservative 0; Mismatches 77; Indels 9; Gaps 2;
 QY 143 ATGGAGTGCATTTACAGTATGGAGTGTCCAGGTGTGAACAACCGTCTGAGCGATATA 202
 Db |||||
 QY 404 AGGGCAGTATTTCACCTGCTGGCAGGCGCTGTGTGAACAACCATGCTCGCGGATACA 345
 Db |||||
 QY 203 GCAAGTGTGGATGCTCAGCTATACATCAGAG-----GGAGGTATGACTTCAGCTACA 256
 Db |||||
 QY 344 GCAATGTGGGTGTTTAAATACGCAACAGCTTCATGGAGGATACAGATTCATGTATC 285
 Db |||||
 QY 257 CTGGACAAACTGCTCTCTCAACACGAGGTGGATGAGTGTGTGCACACACCGAGT 316
 Db |||||
 QY 284 AGGGCCAGACCGCTCGGCTTACACACGACAACTGCAAGGGCGTGTCTCAGACCCGGT 225
 Db |||||
 QY 317 TTGGTCCAGTGCCA---GGGCATGCAACCCCTTTTGGTTGGAAGAGTATCTTCATCCAAT 373
 Db |||||
 QY 224 TTTTTCAGCAGTGTAAATCAAGCTGCACAGTGTGGTTGGAAAAGTTTTTCATCCAGT 165
 Db |||||
 QY 374 GCT 376
 Db |||||
 QY 164 GCT 162
 Db |||||
RESULT 11
 BI416667
LOCUS BI416667
DEFINITION hasp001xn07f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion annosum cDNA clone hasp001xn07f, mRNA sequence.
ACCESSION BI416667
VERSION BI416667.1 GI:15187690
KEYWORDS EST.
SOURCE Pinus sylvestris/Heterobasidion annosum.
ORGANISM Pinus sylvestris/Heterobasidion annosum
 Eukaryota; mixed EST libraries.
REFERENCE 1 (bases 1 to 697)
AUTHORS Asiegbu, F.O., Nahalkova, J., Choi, W., Stenlid, J. and Dean, R.A.

TITLE Expressed sequence tags of randomly selected cDNA clones from the interaction of the root rot fungus (Heterobasidion annosum) with seedling roots of Scots pine (Pinus sylvestris)
JOURNAL Unpublished (2001)
COMMENT Contact: Fred O. Asiegbu
 Dept. of Forest Mycology & Pathology
 Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden
 Tel: +46 18 67 15 98
 Fax: +46 18 30 92 45
 Email: Fred.Asiegbu@mykopat.slu.se
 Seq primer: T7 primer.
FEATURES Location/Qualifiers
 source
 1..697
 /organism="Pinus sylvestris/Heterobasidion annosum"
 /db_xref="taxon:169015"
 /clone="hasp001xn07f"
 /clone_lib="Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp)"
 /dev_stages="Seedling roots of scots pine were infected for 6 days with H. annosum"
 /note="Vector: pT-Adv; Site 1: EcoRI; The subtractive hybridization cDNA library was constructed from scots pine roots infected for 6-days with mycelia of Heterobasidion annosum (FP5)."
BASE COUNT 155 a 145 c 175 g 222 t
ORIGIN
 Query Match 16.6%; Score 81.8; DB 13; Length 697;
 Best Local Similarity 64.6%; Pred. No. 1.4e-12;
 Matches 157; Conservative 0; Mismatches 77; Indels 9; Gaps 2;
 QY 143 ATGGAGTGCATTTACAGTATGGAGTGTCCAGGTGTGAACAACCGTCTGAGCGATATA 202
 Db |||||
 QY 143 AGGGCAGTATTTCACCTGCTGGCAGGCGCTGTGTGAACAACCATGCTCGCAGTACA 202
 Db |||||
 QY 203 GCAAGTGTGGATGCTCAGCTATACATCAGAG-----GGAGGTATGACTTCAGCTACA 256
 Db |||||
 QY 203 GCAAAATGTGGGTGTTTAAATATCGGCAACAGCTTCATGGAGGATACAGATTCATGTATC 262
 Db |||||
 QY 257 CTGGACAAACTGCTCTCTCAACACGAGGTGGATGAGTGTGTGCACACACCGAGT 316
 Db |||||
 QY 263 AGGGCCAGACCGCTCGGCTTACACACGACAACTGCAAGGGCGTGTCTCAGACCCGGT 322
 Db |||||
 QY 317 TTGGTCCAGTGCCA---GGGCATGCAACCCCTTTTGGTTGGAAGAGTATCTTCATCCAAT 373
 Db |||||
 QY 323 TTTTTCAGCAGTGTAAATCAAGCTTGCAGCAGTGTGGTTGGAAAAGTTTTTCATCCAGT 382
 Db |||||
 QY 374 GCT 376
 Db |||||
 QY 383 GCT 385
 Db |||||
RESULT 12
 BI416665
LOCUS BI416665
DEFINITION hasp001xn05f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion annosum cDNA clone hasp001xn05f, mRNA sequence.
ACCESSION BI416665
VERSION BI416665.1 GI:15187688
KEYWORDS EST.
SOURCE Pinus sylvestris/Heterobasidion annosum.
ORGANISM Pinus sylvestris/Heterobasidion annosum
 Eukaryota; mixed EST libraries.
REFERENCE 1 (bases 1 to 531)
AUTHORS Asiegbu, F.O., Nahalkova, J., Choi, W., Stenlid, J. and Dean, R.A.
TITLE Expressed sequence tags of randomly selected cDNA clones from the interaction of the root rot fungus (Heterobasidion annosum) with seedling roots of Scots pine (Pinus sylvestris)
JOURNAL Unpublished (2001)
COMMENT Contact: Fred O. Asiegbu
 Dept. of Forest Mycology & Pathology

Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden

Tel: +46 18 67 15 98
Fax: +46 18 30 92 45

Email: Fred.Asiegbu@mykopat.slu.se
Seq primer: T7 primer.

FEATURES

source

Location/Qualifiers
1..531

/organism="Pinus sylvestris/Heterobasidion annosum"
/db xref="taxon:169015"
/clone="hasp001x17f"
/clone_lib="Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp)"
/dev_stage="Seedling roots of scots pine were infected for 6 days with H. annosum"
/note="Vector: PT-Adv; Site 1: EcoRI; The subtractive hybridization cDNA library was constructed from scots pine roots infected for 6-days with mycelia of Heterobasidion annosum (FPS)."

BASE COUNT 117 a 118 c 138 g 158 t

ORIGIN

Query Match 16.5%; Score 81.2; DB 13; Length 531;

Best Local Similarity 64.3%; Pred. No. 2e-12; Indels 9; Gaps 2;
Matches 157; Conservative 0; Mismatches 78;

143 ATGGAGTGCATTATTAAGATGAGTGTCTCAAGTTGTAAACACCGTGTGACGATATA 202
136 AGGGAGTATTATTCACGTGCGTGGCAGGGCTGTGTAAACACATGCTGCGATACA 195
203 GCAAGTGTGATGCTCACTATACATCAGAG-----GGAGCTATGACTTCACTACA 256
196 GCAAGTGTGATGCTTATATATCGTACACACGTCATGAGGATGAGATTGATATC 255
257 CTGGACAACCTGCTGCTCTACACACGCTGATGCTGTGTGACACACAGCT 316
256 AAGGCCACATCGCTGCGCTTACACACGCGCAACTGCAAGCGCTGCTCATACCCGTT 315
317 TTGGGTCAGTGCCA--GGGCAATGCACACCTTTGGTGAAGATATCTTCATCAAT 373
316 TTCTGCGAGTTATTAACAAGCTTTCGACCGGTTTGGTGGAAAATTTTTCATCACT 375
374 GCTA 377
376 GCTA 379

RESULT 13

B1416726

LOCUS

B1416726 374 bp mRNA linear EST 15-AUG-2001
hasp002xa22f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion annosum cDNA clone hasp002xa22f, mRNA sequence.

DEFINITION

B1416726

VERSION

B1416726.1 GI:15187749
EST.

KEYWORDS

Pinus sylvestris/Heterobasidion annosum.
Pinus sylvestris/Heterobasidion annosum
Eukaryota; mixed EST libraries.

SOURCE

B1416726

REFERENCE

1 (bases 1 to 374)
Asiegbu, F.O., Nahalkova, J., Choi, W., Stenlid, J. and Dean, R.A.
Expressed sequence tags of randomly selected cDNA clones from the interaction of the root rot fungus (Heterobasidion annosum) with seedling roots of Scots pine (Pinus sylvestris)

JOURNAL

Unpublished (2001)
Contact: Fred O. Asiegbu
Dept. of Forest Mycology & Pathology
Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden

AUTHORS

Tel: +46 18 67 15 98
Fax: +46 18 30 92 45

TITLE

Email: Fred.Asiegbu@mykopat.slu.se
Seq primer: T7 primer.

FEATURES

source

Location/Qualifiers
1..374

/organism="Pinus sylvestris/Heterobasidion annosum"
/db xref="taxon:169015"
/clone="hasp002xa22f"
/clone_lib="Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp)"
/dev_stage="Seedling roots of scots pine were infected for 6 days with H. annosum"
/note="Vector: PT-Adv; Site 1: EcoRI; The subtractive hybridization cDNA library was constructed from scots pine roots infected for 6-days with mycelia of Heterobasidion annosum (FPS)."

BASE COUNT 83 a 85 c 104 g 102 t

ORIGIN

Query Match 15.4%; Score 75.8; DB 13; Length 374;
Best Local Similarity 64.6%; Pred. No. 6.2e-11;
Matches 148; Conservative 0; Mismatches 72; Indels 9; Gaps 2;

143 ATGGAGTGCATTATTAAGATGAGTGTCTCAAGTTGTAAACACCGTGTGACGATATA 202
146 AGGGAGTATTATTCACGTGCGTGGCAGGGCTGTGTAAACACATGCTGCGATACA 205
203 GCAAGTGTGATGCTCACTATACATCAGAG-----GGAGCTATGACTTCACTACA 256
206 GCAATGTGGTGTCTTATATCGCAACACGTTCAATGAGGATGAGATTGATATC 265
257 CTGGACAACCTGCTGCTCTACACACGCTGATGCTGTGTGACACACAGCT 316
266 AAGGCCACATCGCTTGTCTTACACACGCAACTGCAAGGGCTGCTCAGACCCGCT 325
317 TTGGGTCAGTGCCA--GGGCAATGCACACCTTTGGTGAAGATAT 362
326 TTCTGAGAGTTATTAACAAGCTTTCGACCGGTTTGGTGGAAAATTTTTCATCACT 374

RESULT 14

B1416558/c

LOCUS

B1416558 476 bp mRNA linear EST 15-AUG-2001
hasp001xf17f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion annosum cDNA clone hasp001xf17f, mRNA sequence.

DEFINITION

B1416558

VERSION

B1416558.1 GI:15187581
EST.

KEYWORDS

Pinus sylvestris/Heterobasidion annosum.
Pinus sylvestris/Heterobasidion annosum
Eukaryota; mixed EST libraries.

SOURCE

B1416558

REFERENCE

1 (bases 1 to 476)
Asiegbu, F.O., Nahalkova, J., Choi, W., Stenlid, J. and Dean, R.A.
Expressed sequence tags of randomly selected cDNA clones from the interaction of the root rot fungus (Heterobasidion annosum) with seedling roots of Scots pine (Pinus sylvestris)

JOURNAL

Unpublished (2001)
Contact: Fred O. Asiegbu
Dept. of Forest Mycology & Pathology
Swedish University of Agriculture, Box 7026, S-750 07, Uppsala, Sweden

AUTHORS

Tel: +46 18 67 15 98
Fax: +46 18 30 92 45

TITLE

Email: Fred.Asiegbu@mykopat.slu.se
Seq primer: T7 primer.

FEATURES

source

Location/Qualifiers
1..476

/organism="Pinus sylvestris/Heterobasidion annosum"
/db xref="taxon:169015"
/clone="hasp001xf17f"
/clone_lib="Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp)"
/dev_stage="Seedling roots of scots pine were infected for 6 days with H. annosum"
/note="Vector: PT-Adv; Site 1: EcoRI; The subtractive

hybridization cDNA library was constructed from Scots pine roots infected for 6-days with mycelia of *Heterobasidion annosum* (FP5)."

BASE COUNT
ORIGIN

141 a 114 c 91 g 129 t 1 others

Query Match 11.6%; Score 57.4; DB 13; Length 476;
Best Local Similarity 66.7%; Pred. No. 1.3e-05;
Matches 98; Conservative 0; Mismatches 46; Indels 3; Gaps 1;

QY 233 AGGAGGCTATGACCTTACAGTACACTGACAACTGCTCTCTACAAACAGGCTGGAT 292
DB 452 ATGGAGGATACAGTTCTATGATCAAGCCAGACCGCTGCGGCTTACAAACGAGCAACT 393
QY 293 GCAGTGGTGTGCACACACACAGGTTGGGTCAGTGCAGG---GCATGCAACCCCTTTTG 349
DB 392 GCAAGGGCGTTGCTCAGACCCGGTTTCTAGCAGTGTGTTAATCAAGCTTGACGACGATTTTG 333
QY 350 GTTGAAGAGTATCTTCATCCATGCT 376
DB 332 GTTGAAGAGTTTTTTTCATCCAGTCT 306

RESULT 15

BE662406
LOCUS 294 bp mRNA linear EST 08-SEP-2000
DEFINITION ST62/ST62H06 Pine Triplex shoot tip library Pinus taeda cDNA clone
ST62/ST62H06, mRNA sequence.

ACCESSION BE662406

VERSION BE662406.1 GI:9992999

KEYWORDS EST.

SOURCE loblolly pine.

ORGANISM Pinus taeda

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.

1 (bases 1 to 294)

Whetten, R.W., Kinlaw, C.S., Retzel, E. and Sederoff, R.R.

The Pine Gene Discovery Project

Unpublished (1999)

Contact: Ross Whetten

Forest Biotechnology Group

North Carolina State University

Dept. of Forestry, NC State University, 6113 Jordan Hall, Raleigh

, NC, 27695-8008

Tel: 919-515-7800

Fax: 919-515-7801

Email: rosswhetten@unity.ncsu.edu

Seq primer: 5' lambda Triplex2 Sequencing Primer.

Location/Qualifiers

1. .294

/organism="Pinus taeda"

/db_xref="taxon:3352"

/clone="ST62/ST62H06"

/clone_lib="Pine Triplex shoot tip library"

/lab_host="E. coli BM25.8"

/note="Organ: shoot tips; Vector: Lambda Triplex; Site: 1:
SfiI (A); Site 2: SfiI (B); Shoot tips (approx. 2 cm from
apex) were collected during the spring, frozen and used
for mRNA isolation. The SMART-PCR method (Clontech) was
used to prepare a library from 1 ug total RNA, using the
Lambda Triplex vector. Plasmid subclones in pTriplex were
recovered by cre-lox excision in E. coli strain BM25.8 and
sequenced from the 5' end."

BASE COUNT 70 a 67 c 69 g 77 t 11 others

ORIGIN

Query Match 11.6%; Score 57; DB 10; Length 294;
Best Local Similarity 60.2%; Pred. No. 1.5e-05;
Matches 106; Conservative 0; Mismatches 64; Indels 6; Gaps 1;

QY 143 ATGGAGTGCATTTACAGTATGAGTGTCCAGGTTGTAAACACCGCTGAGCGATATA 202
DB 118 AGGGCANTATTCTACTCGGTGGGAGGCGCTNNNTTGTAAACACCATGCTGCGGATACA 177

QY 203 GCAAGTGTGGATGCTCAGCTATACATCAGAAG-----GGAGGCTATGACTTCAGCTACA 256
DB 178 GCAAAATGTGGTGTCTTAATATCGGCAACAACGTTTCATGGAGGATACGAGTTTCATGTATC 237
QY 257 CTGGACAAACTGCTGCTCTCTACAAACCGGCTGGGATGCAAGTGGTGGTGGCACACACC 312
DB 238 AAGNNCAGACCCNNNTGCGNNNTTAAACACCGGCAACTGCAAGGGCGTTTGTCAAACC 293

Search completed: January 12, 2003, 08:35:41
Job time : 1953 secs